Clustering and Shifting of Regional Appearance for Deformable Model Segmentation

Joshua V. Stough

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill
2008

Approved by:

Stephen M. Pizer, Ph.D., Advisor

Edward L. Chaney, Ph.D., Co-principal Reader

J. S. Marron, Ph.D., Reader

Mark Foskey, Ph.D., Reader

Marc Niethammer, Ph.D., Reader

Abstract

Joshua V. Stough: Clustering and Shifting of Regional Appearance for Deformable Model Segmentation
(Under the direction of Stephen M. Pizer, Ph.D.)

Automated medical image segmentation is a challenging task that benefits from the use of effective image appearance models. An appearance model describes the grey-level intensity information relative to the object being segmented. Previous models that compare the target against a single template image or that assume a very small-scale correspondence fail to capture the variability seen in the target cases. In this dissertation I present novel appearance models to address these deficiencies, and I show their efficacy in segmentation via deformable models.

The models developed here use clustering and shifting of the object-relative appearance to capture the true variability in appearance. They all learn their parameters from training sets of previously-segmented images. The first model uses clustering on cross-boundary intensity profiles in the training set to determine profile types, and then it builds a template of optimal types that reflects the various edge characteristics seen around the boundary. The second model uses clustering on local regional image descriptors to determine large-scale regions relative to the boundary. The method then partitions the object boundary according to region type and captures the intensity variability per region type. The third and fourth models allow shifting of the image model on the boundary to reflect knowledge of the variable regional conformations seen in training.

I evaluate the appearance models by considering their efficacy in segmentation of the kidney, bladder, and prostate in abdominal and male pelvis CT. I compare the automatically generated segmentations using these models against expert manual segmentations of the target cases and against automatically generated segmentations using previous models.

Acknowledgments

I would like to thank many people who have been important to this work and more generally to my growth while here at UNC.

To start, Steve Pizer has been critical. His ideas are integral to this work and his feedback and quick response to edits greatly sped up the writing process. Moreover, I am both happy and mortified to see myself having borrowed many of his quirks, including an obsession with precision in speech.

Ed Chaney's knowledge and enthusiasm has kept me excited about the research through all the years. Steve Marron is a great teacher of statistics who has always either been able to explain ideas to me or to know where to find a good explanation, and his Matlab scripts for statistical analysis have been incredibly helpful. Mark Foskey has been a mentor and is a good ultimate frisbee player to boot. I only just met Marc Niethammer, but I am appreciative of his extensive feedback and ideas. Other faculty who have helped along the way (many as members of my committee) are Surajit Ray, Sarang Joshi, Guido Gerig, and Keith Muller.

The people in the MIDAG group have been great. Graham Gash, Gregg Tracton, Josh Levy (pipeline), and Eli Broadhurst (RIQFs) have all been critical in this research. Tom Fletcher, Ja-yeon Jeong, Derek Merck (hilarious), Xiaoxiao Liu, and Rohit Saboo, and others.

More related to my time here than to the dissertation itself, I would like to thank the team of people who run the department. I have relied heavily on their knowledge and friendliness. Thank you Tammy and Donna for knowing who to ask: thank you Mike and Charles for the excessive technical support. I would also like to thank those responsible for allowing me to teach an extra course, who I believe are Pizer, Steve Weiss, David Stotts, and Tim Quigg, in no particular order.

I am grateful for my wife Mia. She is the most clever person I know. Her ability to help me organize my thoughts have been very important for my writing, whether on the dissertation or on my job applications (research/teaching statements, cover letter). And as a Ph.D. in

genetics, she has taught me all about end-to-end chromosome fusions in C. elegans, which makes me a slightly more interesting person at parties.

My parents have been waiting for me to get a real job for quite a while. I am thankful for their patience.

Thomas Hobbes: True and false are attributes of speech, not of things.

Philander Chase Johnson: Cheer up! The worst is yet to come!

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Medical Image Segmentation

In this dissertation I present novel image appearance models and their application in automatic medical image segmentation. Medical image segmentation refers to the task of delineating objects such as organs in 2D and 3D medical images (see Figure 1.1 for an example). High quality segmentation is a critical aspect of many medical applications, both clinical and research oriented, including surgical planning and image-guided surgery, radiation therapy, diagnosis, and analysis of the relationship between anatomic geometry and disease. The "gold standard" in segmentation is that performed by trained expert clinicians, for whom it is time intensive and thus high cost. Additionally, there may be important discrepancies in delineations of the same target for different clinicians and even the same clinician at different times.

As a means of improving segmentation precision and cost in the above applications, automating medical image segmentation is a goal common to many researchers around the world. Several paradigms have arisen to this end, including pixel- or voxel-scale decision methods, where an inside/outside decision is made for every image element, and image warping methods, where a target image is registered to a previously expert-segmented image and the previous case's segmentation is transferred. For objects that

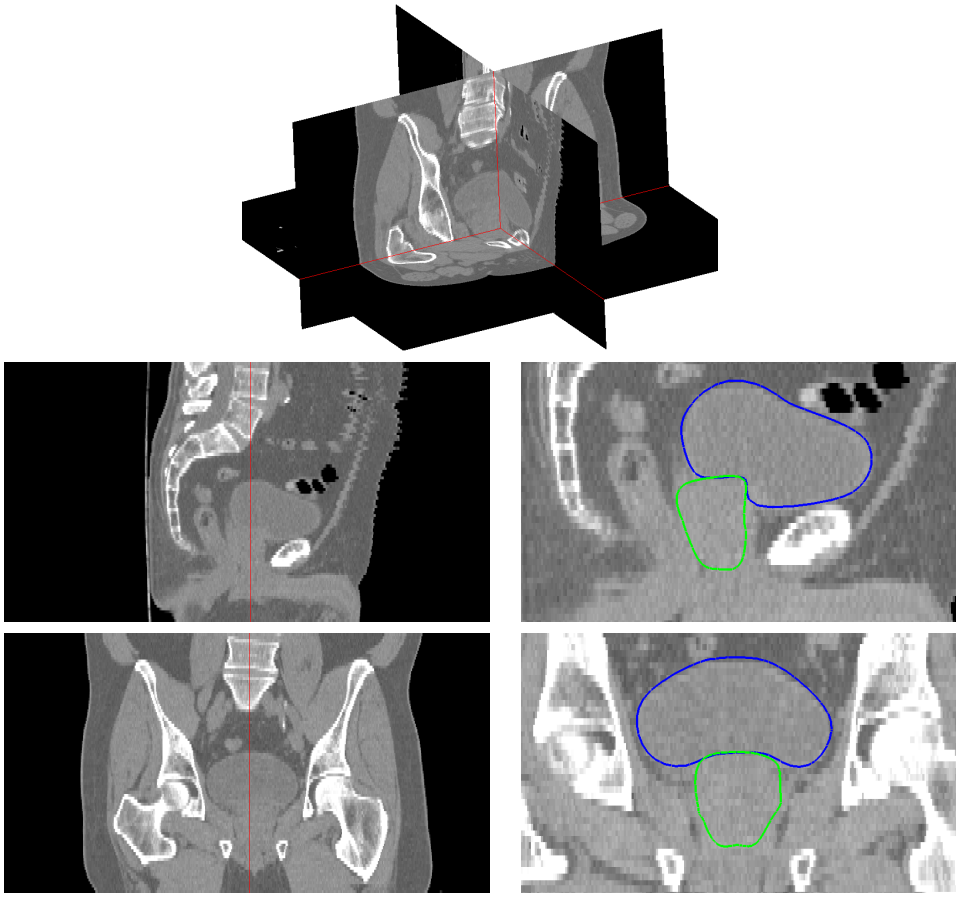

Figure 1.1: Top: Orthogonal slices of a 3D computed tomography (CT) image of the male pelvis, shown off-axis. Middle: Sagittal slice with bladder (blue) and prostate (green) segmentations. Bottom: Coronal view with corresponding segmentations. The red lines show the intersection of two orthogonal slices.

can be described as having simple topology, another class of methods that has proven highly successful is termed deformable model segmentation (DMS), where some geometric model is deformed to fit an image so that the deformed model is consistent with evidence for the object in the image. A subclass called Bayesian deformable model methods is characterized by the use in optimization of (sometimes nominally) probabilistic terms that reflect the shape and appearance of the object of interest.

All automatic segmentation schemes depend on appearance models, which describe the intensity information in an image or in some anatomy-relative part of an image. Within the Bayesian DMS context, some appearance models found elsewhere capture

appearance in an object-relative and regional way. I have developed novel appearance models that extend such models and that are characterized by clustering and/or shifting on object-relative regional image descriptors. To justify *object-relative* appearance, consider that the shape model that fits the organ of interest may define a coordinate system on the image that is object-relative. This coordinate system has the advantage that anatomical relationships between the organ of interest and its neighbors are described more naturally than by using the simple voxel coordinates. Such anatomical relationships may vary from person to person and day to day. To justify *regional* appearance, consider that an image can be decomposed into regions such as organs that are distinct from one another in their intensity characteristics. With these keys in mind, my appearance models for segmenting an object in an image use *clustering* to determine these regions and their intensity patterns and *shifting* to account for the variability in the object-relative position of these regions.

The remainder of this chapter is organized as follows. Section 1.2 reviews Bayesian deformable model segmentation and the objective function underlying such optimizations. Section 1.3 describes image appearance models and posits the advantages of object-relative and regional models. The chapter concludes with the thesis statement, a list of accomplishments toward this thesis, and an overview of the remaining chapters of this dissertation.

1.2 Bayesian Deformable Model Segmentation

In Bayesian deformable model segmentation for medical images, a geometric model for an object of interest is initialized in an image and then deformed via its shape parameters m to fit an image object that the model represents. These methods then search for the deformed model (i.e., the prospective segmentation) that optimizes some objective function on m and the image data I. In one particular formulation, this objective

function is the posterior probability of the shape model parameters given the image data, $\arg\max_m p(m|I)$.

Informally, consider an average shape of an object and some target image that includes a particular instance of that object. Bayesian deformable model methods then look to deform the average shape as little as possible while fitting the particular object seen in that image as well as possible. The keys to this process are how to compute an average shape and measure the size of a deformation, which are aspects of shape, and how to measure the fit of a deformed model to the particular image object, which is an aspect of image appearance. These measurements are made using probability distributions on shape and image appearance respectively. Often, the probability distributions are learned from a training set of previously-segmented cases of the same object and image modality.

The keys to the Bayesian process can be understood by applying Bayes' formula to the posterior:

$$p(m|I) = \frac{p(m, I)}{p(I)} = \frac{p(m) \cdot p(I|m)}{p(I)}. \tag{1.1}$$

In that $p(I)$ is constant for any given image, the posterior $p(m|I)$ is proportional to $p(m) \cdot p(I|m)$, and thus

$$\arg\max_m p(m|I) = \arg\max_m [p(m) \cdot p(I|m)]. \tag{1.2}$$

This decomposition leads to segmentations that proceed based on the optimization of an objective function that includes terms measuring two aspects of the model fit to the image object. One aspect is called geometric typicality, and it measures the prior probability on the shape parameters, $p(m)$. The other aspect is called the geometry-to-image match–*image match* for short–and is given by the likelihood of the shape parameters given the image data, or $p(I|m)$. The optimization can be reformulated as a minimization by applying the negative log;

$$\arg\min_m \left[-\log p(m|I)\right] = \arg\min_m \left[-\log p(m) - \log p(I|m)\right]. \tag{1.3}$$

In the above equation the two terms $-\log p(m)$ and $-\log p(I|m)$ are penalties against low probability in the shape and image distributions respectively. This formulation is often used given the plethora of minimization schemes such as the gradient descent and conjugate gradient methods (Shewchuk, 1994). Additionally, in the case that the prior and likelihood are each given by a Gaussian model, the negative log cancels the exponential of the Gaussian and the result is the square Mahalanobis distance (Mahalanobis, 1936), a popular and efficiently calculated statistical distance metric well-suited – in that it is quadratic – to the these minimization techniques.

In practice, the prior and likelihood probability distributions are often defined heuristically. Ideally however, they are trained given a set of previously expert-segmented cases in order to drive the optimization to expert-like results.

This dissertation focuses on the image match term and its underlying appearance models. The following section describes image match in more detail and presents male pelvis segmentation in computed tomography (CT) as motivation. Following that, I present my thesis statement and claims, and an outline of the rest of the dissertation.

1.3 Image Match and the Image Appearance Models

The image match term defines the fit of the deformed object to the image data; it is a result of the appearance model. An appearance model describes the grey-level intensity information relative to the object being segmented, typically relative to the object's boundary.

Consider, for example, the segmentation problem of finding a disk in a 2D image. If

the intensity within the disk is brighter than outside the disk, an effective appearance model would capture that information. Initial attempts considered the edge strength (gradient magnitude) of the image at the boundary of prospective disks. However, several image aspects common to medical segmentation make such a model inadequate. First, the presence of noise in the image misinforms the segmentation process as to the boundary of the disk. Adding to this is the fact that the entire boundary of the object in an image is not always characterized by a change in intensity. A better model might consider the average intensities inside and outside the disk. The average is an example of a *parameterization* of the intensity information taken with some aperture (here, the inside or outside). A resulting image match could be the difference between these inside and outside averages, and a prospective disk that maximizes this difference ($\bar{i}_{inside} - \bar{i}_{outside}$) would be a good segmentation. A problem with this match could be that a collapsed disk containing only the brightest pixel in the image would maximize the inside/outside difference. However, recall that the shape prior would prevent this from happening were such a disk unlikely.

For a *probabilistic* image match in the same context, one could capture the average inside and outside intensities for a set of training images and fit some distribution (e.g., Gaussian) to them. The resulting image match could then be the joint probability of the inside and outside averages, i.e. $p(I|m) = p(\bar{i}_{inside}, \bar{i}_{outside})$, where the prospective disk's shape parameters m (e.g., center and radius) define inside and outside. One subtlety here that is relevant throughout this work is that I have substituted a probability of I *given* m with a probability of I *relative to* m. This assumes that regions in the same position relative to an object consist of the same image intensities regardless of the object's current deformation. In that intensity is the result of tissue composition, which does not change with deformation, this is a reasonable assumption for my target organs.

The complexity of the appearance model must be commensurate with the complexity of the image problem. Figure 1.1 shows that a simple light-on-dark appearance model

as envisioned above will not be adequate for some applications such as pelvic organ segmentation in CT. Often a target object may have lighter and darker parts within itself, so a single number such as average intensity does not adequately characterize its appearance or the variation in appearance across images.

In the work I present here, I use two parameterizations that more completely describe the intensity information. The first is termed profiles. A profile is a tuple of intensity sampled from inside to outside the object along the normal to a specified point on the boundary; profiles are obtained for many points on a single object boundary. A profile provides a very local description of the appearance across the object boundary at a point. The second parameterization, called a *quantile function*, records a function of the empirical distribution on intensity within a region. This function of the regional histogram has the attractive property that certain common changes in a distribution are represented linearly in the feature space (Broadhurst et al., 2006). Due to this linearity, quantile functions are amenable to Euclidean statistical methods (e.g., principal component analysis) for modeling the variation in appearance across images.

1.3.1 Object-relative, regional appearance

This subsection motivates the use of object-relative and regional image appearance models.

The purpose of the appearance model is to characterize the intensity pattern of the some target object or region across images. For medical images, this variability comes in several forms. The image including a particular target region may vary across days and patients according to the region's surroundings and tissue density or other physical property that is manifested in intensity. (This is in addition to changes in shape.) A redeeming quality of the variability is that the appearance is more consistent across target cases when considered in an *object-relative* and *regional* way, more than just inside and outside. The example images in Figure 1.2 show that over the target images

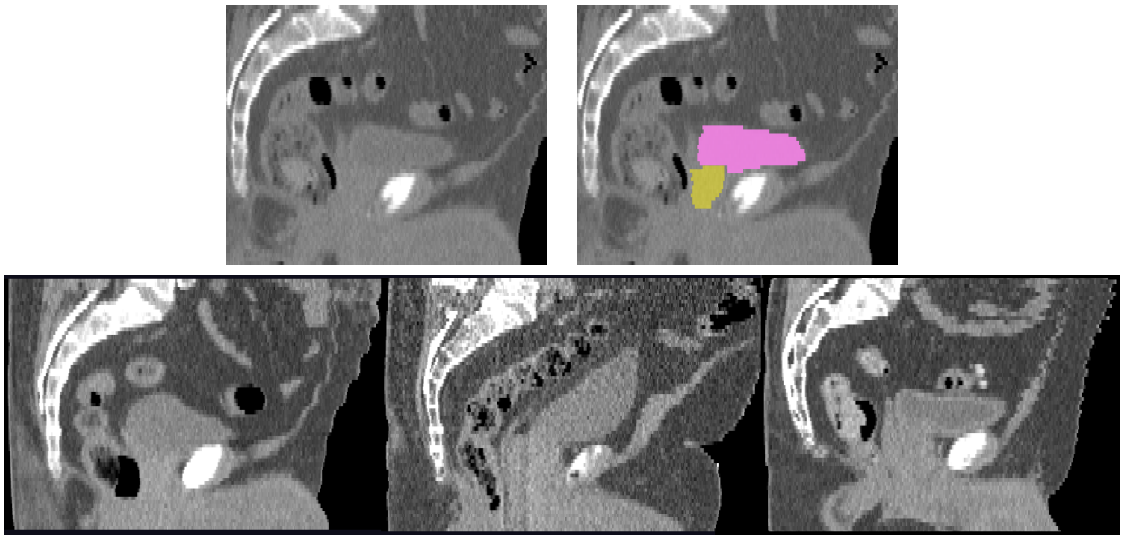

Figure 1.2: Sagittal male pelvis slices in CT. Top: Original greyscale image on the left, with bladder (pink) and prostate (gold) highlighted on the right. Bottom: Across images, note that the bladder is surrounded by mostly lower intensity bowel and fat, with much brighter tissue from the pubic bones and prostate inferior to (below) it. The prostate has a similarly consistent object-relative appearance. These observations motivate an object-relative regional appearance model.

there can be regions of similar intensity at corresponding places relative to the object geometry.

Consider the prostate as an example. A pelvic image including the prostate exists in a space in which the coordinates of a voxel are relative to the scanner. If the patient is imaged twice in differing positions relative to the scanner, then at the voxel coordinates where there is prostate in one image there may be no prostate in the other, and the images are completely different. This is because voxels with the same scanner-relative coordinates are assumed to be in correspondence. However, the images *are* identical if the voxels are sampled more naturally according to a coordinate system implied by the prostate geometry. For example, 2 cm superior to the prostate (toward the head) is 2 cm superior to the prostate in both images and will have about the same intensity.

To continue with the example, an appearance model can take advantage of this *object-relative* correspondence. The bladder, and its appearance, is always superior to

the prostate, while the pubis bones, and their appearance, are anterior to (in front of) it. An effective appearance model in this situation will account for these differing expected intensities relative to the object. Determining the correspondence required for such an object-relative appearance model is a much larger problem than is completely addressed in this work (see (Davies et al., 2003; Cates et al., 2007) for a larger discussion on determining correspondence). For the appearance models described here, the initial correspondence is provided by the deformable model geometry, discussed later. For two models, that geometric correspondence is relaxed through shifting of the appearance model on the surface of the object model (see below).

A disadvantage of previous models is the use of "global" inside/outside regions. Global regions ignore the natural decomposition of the image space, especially in the exterior of the object of interest. Given the subject (a human body), medical images are populated by regions of relative homogeneity in tissue composition, which is reflected by homogeneity in the local intensity. Consider that what constitutes such a region is an organ or other volume whose local intensity distributions are distinguishable from those of neighboring volumes. In the example above the prostate is surrounded by regions of homogeneity that represent the bladder and pubis bones, rectum and neighboring fat deposits. These regions are a cause of intensity inhomogeneity in the organ exterior.

In this work I find regions through *clustering* on local intensity distributions. To avoid the infinity of possible regional shapes and positions, the regions are constrained to be on the boundary of the object of interest and extended normal to the boundary, inside or outside.

The variability in the positions of these regions is also lost on global appearance models. At the same time, the sliding of neighboring regions with respect to one another and the object of interest cause problems for methods that assume a voxel scale correspondence (see Chapter 2). Two of the image matches I propose in this work use *shifting* of the appearance model on the boundary to account for the variable object-relative

positions of these regions.

1.4 Thesis

Thesis: *The automatic segmentation of medical images via deformable models benefits from the use of object-relative, regional statistical appearance models. Specifically, clustering on either local or regional object-relative image descriptors can provide an effective understanding of the context of the object of interest in the image. Additionally, the use of shifting to account for the variability in the position of object-relative regions allows for relaxation of the geometric correspondence provided by the deformable model and often leads to improved segmentations.*

1.5 Claims

Contributions included in this dissertation are as follows:

1. I demonstrate that clustering on local image descriptors can be used to construct effective appearance models for deformable model segmentation.

2. I present a novel appearance model and image match using clustering on intensity profiles that reflects the various edge characteristics in the image boundary region. This model leads to improved kidney segmentation in CT versus a previous image match.

3. I demonstrate that regional intensity quantile functions (RIQFs) are a more effective local image descriptor than intensity profiles.

4. I present a novel appearance model and image match using clustering on RIQFs. I compare RIQF-based appearance models at several regional scales on bladder

and prostate segmentation in CT.

5. I demonstrate that the object-relative positions of neighboring organs and volumes change across images, leading to a false association of image regions due to the geometry-implied correspondence.

6. I present two novel appearance models and associated image matches that account for this variability in the external conformation of neighboring regions through shifting of the image model on the object boundary. The two models differ in the scale of the regions undergoing shifting. These models often lead to improved segmentations over those achieved using their static counterparts. I validate this by experimentally testing the efficacy in segmentation of these appearance models on bladder and prostate in CT.

1.6 Overview of the Chapters

This dissertation is organized in six chapters, with this chapter introducing and motivating the work. The remainder of the dissertation is organized in the following:

Chapter 2 provides a review of previous image matches and Bayesian deformable model segmentation in general. Also presented is an introduction to the deformable model of choice in this work, the discrete medial representation, or m-rep. Finally, this chapter reviews in more depth the image descriptors used in this work, intensity profiles and regional intensity quantile functions.

Chapter 3 details an image appearance model and image match based on clustering of object-relative intensity profiles. This includes a means of clustering on intensity profiles using a correlation metric, construction of an image match that is optimal in this metric, and validation of the model on kidney segmentation in CT using the m-rep deformable model segmentation scheme.

Chapter 4 presents a comparison of appearance models at three object-relative regional scales from global to local. This chapter also presents an appearance model and image match based on a static decomposition of the object boundary and clustering on local RIQFs. After determining region types through clustering, the method partitions the object boundary according to region type and captures the intensity variability per region type using principal component analysis (PCA). This model's efficacy is evaluated on bladder and prostate segmentation in CT.

Chapter 5 presents two appearance models and their associated image matchs that improve upon the models of Chapter 4 by allowing the region-type partition on the boundary to slide according to variability in the partition seen in training. The changing in the partition is accomplished using an iterative process. The image matches are also validated in this chapter, on bladder and prostate segmentation in CT.

Chapter 6 concludes with a review of the contributions of this dissertation, a discussion of the results of the research, and possible future directions for research.

Chapter 2

Background

Appearance models and the resultant image matches are an integral part of all automatic segmentation schemes. They define the fit of a prospective segmentation to the image data and thus are required for any optimization to take place. The appearance models I have developed in this dissertation can be better understood within the context of previous models. In this chapter I present a survey of appearance models and discuss their strengths and weaknesses.

I classify appearance models according to the scale of the image match computation. *Voxel-scale* models consider the image information at the level of individual ordered voxels[1]. *Region-scale* models consider information over many voxels at once (e.g., the intensity distribution in the interior of the object of interest). Within these two categories, I further classify appearance models according to whether the model encodes variability in appearance: models are either *static* or *probabilistic*.

[1]While I usually refer to the image information as intensity, which is most common in the literature I cite, this information can go beyond a single intensity per voxel, as in multispectral or diffusion tensor imaging. In these imaging modalities operators like the gradient must be made appropriate to the intensity space.

2.1 Voxel-scale Models

The simplest voxel-scale model assumes that the object edge corresponds to a local maximum in the image gradient magnitude (Canny, 1986). The active contour and "snake" segmentation methods of (Kass et al., 1988; Caselles et al., 1997) both use this idea, attenuating a local growth force on the shape model according to a function of the local gradient such as magnitude. The overall image match is then proportional to the integral of this local match over the whole contour S,

$$p(I|m) \propto \oint f(\nabla I(S(u)))du. \tag{2.1}$$

(Cootes et al., 1995) also use an edge strength image match within the context of the notable Active Shape Model segmentation scheme (see below). Here, the authors use the derivative of the intensity profile at an object boundary point to determine the displacement of that point along its normal. In the work of (Staib and Duncan, 1992), the edge strength image match is paired with the Fourier-parameterized boundary model.

The gradient magnitude image match works on the assumptions that the boundary of the object in the image is characterized by a relatively sharp change in intensity and that this change is common over the entire boundary. For organs and imaging modalities where these assumptions hold, the gradient magnitude match can lead to impressive results, though often requiring user guidance (Yushkevich et al., 2006). However, due to the nature of medical images, such assumptions are often inadequate. Noise in an image may create new edges that are stronger than the edge of the object. Furthermore, nearby organs, bones, and other tissue lead to different appearance in different parts of the boundary. The physical characteristics of the object tissue and neighboring tissue leads to less contrast in some parts of the boundary than in others.

2.1.1 Static template methods

Several approaches have been proposed to account for the differing appearance around the object. Some methods use the appearance from a single example image to create a static template to match against at target time. Examples of this are found in the medially-based deformable model segmentations of (Pizer et al., 2003) and in the diffeomorphic image warping of (Miller et al., 1999).

In (Pizer et al., 2003; Joshi et al., 2001), the medial representation, or *m-rep* shape model is used within the Bayesian DMS framework (see Chapter 1). In this scheme, intensity profiles are sampled for a collar region about the boundary of the shape that fits a single expert-segmented image. The collection of profiles is termed a simple mask, in that it is trained on only one image. At target time, the corresponding profiles are sampled for the prospective shape in the new image. The image match is then normalized correlation of the target profiles with the profiles from the training case: $p(I|m) \propto X \cdot T$, where X is a column vector formed by the concatenation of the target profiles and T is that from the training case. In an approach that mimics the gradient magnitude image match above, the training profiles can be substituted with profiles that encode an intensity step edge from inside to outside, implying a bright object on a dark background or vice versa. An improved image match using the profile descriptor is the subject of Chapter 3. The m-rep shape model is covered later within this chapter (see Section 2.4), as my image matches also use the object-relative correspondence it provides.

Another commonly used voxel-scale, static template image match is the sum of squared differences match in (Miller et al., 1999). In this scheme a diffeomorphic warp according the Navier-Stokes fluid model is computed to deform the template image to fit the target image. Given a warp, structures of the static template such as associated segmentations can be transferred to the target according to the warp. The image match of a prospective warp is the sum over the image of the square difference between

corresponding voxel intensities in the deformed template and the target,

$$p(I|m) \propto \oint_{\Omega} |T(x - u(x)) - S(x)|^2 dx, \tag{2.2}$$

where u is a displacement field, T is the template image and S is the target. Davis et al. use this scheme to perform regression on ordered images and construct a moving average image, such as the aging brain in MRI (Davis et al., 2007; Davis, 2008). Regional and probabilistic image matches in the image registration literature are covered below.

Mutual information is another image match popular particularly for the registration of images of different modalities (Maes et al., 1997). Here, as in the above, the purpose is to determine a warp (here constrained to be affine) that deforms the template image to fit the target. The mutual information (MI) image match is entropy based, essentially measuring the degree to which the two images are *not* independent; that is, how well a voxel in the deformed template image predicts the value of the corresponding voxel in the target. Seen one way, MI measures the difference between the sum of the individual entropies of the two images, seen as random variables X and Y, and the joint entropy:

$$MI(X, Y) = H(X) + H(Y) - H(X, Y). \tag{2.3}$$

For another way to think of mutual information, consider that $p_X(x)$ and $p_Y(y)$ are the marginal distributions of X and Y and $p_{X,Y}(x, y)$ is the joint distribution, all as histograms for example. If X and Y are independent, then the joint distribution is equal to the product of the marginals: $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$. MI then measures the degree to which the product of the marginals is *not* the joint, using the Kullback-Leibler divergence between the joint and the product of the marginals,

$$MI(X, Y) = \sum_{x,y} p_{X,Y}(x, y) log \frac{p_{X,Y}(x, y)}{p_X(x) \cdot p_Y(y)}. \tag{2.4}$$

(Studholme et al., 1999) construct a normalized MI measure, the ratio of the sum of the individual entropies and the joint, that they claim improves the robustness of registration under certain conditions. (Gan and Chung, 2005) also claim improved registration using MI on a feature vector including both the traditional intensity and what amounts to a smoothed gradient magnitude. (Tsai et al., 2004) propose a mutual information-based active contour segmentation where the image match is the mutual information between the target image intensities and the prospective segmentation labels, as opposed to the template image intensities.

2.1.2 Probabilistic methods

The static template models were devised to account for the differing appearance around the object. However, being trained on a single example case, they capture limited information. Among sets of images, there is additional variability even among anatomically corresponding voxels. Several probabilistic, still voxel-scale models have been proposed to account for this variability. These probabilistic appearance models lead to optima with the most statistically *likely* image characteristics rather than optima *most like* a single arbitrarily chosen training case.

Chief among these schemes is the well-known Active Shape Models of Cootes et al. (1994). This approach is also profile-based, where in training, intensity profiles are sampled normal to the object boundary at corresponding places in the training images. As opposed to the m-rep based correspondence of Pizer et al. (2003), here correspondence is provided by an underlying Point Distribution Model (PDM). The normalized derivatives of the profiles observed at a corresponding point across the training cases form a population on which to train the appearance at that point. The authors use Principal Component Analysis (PCA, see Section 2.3.2) to construct probability distributions on these feature vectors at each point. At target time then, the corresponding profiles are sampled and the image match at a point is the square Mahalanobis distance of the

target profile in its respective distribution. Kelemen et al. use this PCA-per-point appearance model on the original grey-scale intensity profiles. Their work is in the context of deformable model segmentation with the spherical harmonics shape model (Kelemen et al., 1999).

Many statistical appearance models assume for practicality that the elements of a feature vector such as a profile are independent. This assumption may not be ideal, in that the samples forming the profile are highly spatially correlated and thus some correlation in the intensity can be expected. Duta et al. (1999) present another probabilistic, profile-based appearance model, but in this case partially accounting for the correlation by modeling the profile as a Markov chain. The purpose in this case is to detect places in the target image that are statistically likely to be the object of interest, given training. The scheme involves sampling a small number of intensity profiles at a fixed number of positions through a square region of the image. The concatenated profiles form the feature vector for that region. The feature vectors sampled over all positions within all training images form two populations, those vectors corresponding to positions that were in the object of interest and positions that were not; positive and negative examples respectively. The method then determines a maximally discriminative Markov chain model for the feature vectors of the two populations. At target time, positions are classified as the object of interest if the log-likelihood ratio exceeds some threshold T:

$$L(x) = \log \frac{P(x)}{N(x)} > T > 0, \tag{2.5}$$

where x is the position's profile in the target image–the target feature vector–and P and N are probability distributions for the positive and negative populations respectively.

Fenster and Kender (2001) propose a probabilistic appearance model within the snake segmentation framework, called "sectored snakes," that takes into account the differing appearance around the object of interest. In this scheme, the feature vector at a point on the boundary consists of the intensity at that point and the image derivative in

the direction of the normal to the boundary at that point. The method considers the set of feature vectors sampled at a corresponding point across training. In that the original active contour methods do not require correspondence–the edge-strength model is the same regardless of position along the boundary–the correspondence here is provided by the user and by preprocessing alignment steps. The intensity and directional derivative of the feature vector are each modeled with a Gaussian distribution. At target time the image match of a prospective segmentation is the Mahalanobis distance of the target feature vector, integrated over the contour S. That is, $p(I|m) \propto e^{-E_I}$, where

$$E_I = \oint \frac{(I(S(u)) - \mu_I)^2}{\sigma_I^2} du + \oint \frac{(S^{\perp}(u) \cdot \nabla I(S(u)) - \mu_{\nabla})^2}{\sigma_{\nabla}^2} du \tag{2.6}$$

While many of the appearance models presented to this point consider the intensity or some function of intensity or the variability in such, these appearance models consider these characteristics only within a limited region about the boundary of the object of interest–a collar region. However, when almost all parts of an image contain information that can help localize the object (or when, as in registration, the alignment of the image itself is the goal), the scope of the appearance model should encompass the entire image. Cootes et al. propose such a model in their Active Appearance Models (AAM) (Cootes et al., 2001). In this approach the shape model is a PDM of salient points spread throughout the image; the "shape" for an image is the ordered list of these points. A warp defines the alignment of each training shape to the average shape, and this warp is used to align the training images and thus place all their pixels in correspondence. The feature vector is then the entire normalized and aligned image as a column vector. The appearance is obtained using PCA on the population of feature vectors over the training images. A hierarchical appearance results from this training at multiple blurring levels in the training images. At target time, the segmentation procedure iterates from more blurred to less blurred appearance to improve robustness.

For a particular blurring level, the goal is to model the target image in terms of the warping parameters t to the average shape, the coefficients c on the modes of the PCA model on intensity, and intensity normalization parameters u. The image match is then simply the sum of squared differences between the actual target image and the modeled target. The authors further devise a scheme for analytically determining the change in $p^T = (c^T|t^T|u^T)$ required to improve this image match. (Scott et al., 2003) extend the AAM framework by considering not only the intensities but also image filter responses that encode local texture and geometric information, replacing a single intensity at each pixel with a "local structure tuple."

There is additional segmentation literature focused on classifying each voxel in an image (Gerig et al., 1992; Priebe et al., 2006; McLachlan et al., 1996; Woolrich et al., 2005; Duda et al., 2001; Jain et al., 2000). These schemes use training cases to model the classes of tissue being segmented, usually using Gaussian or mixture of Gaussian distributions. Though they also often involve the use of regularizing and spatial prior techniques to limit the noise in the classification, they can still be considered voxel-scale.

Voxel-scale models are characterized by an understanding of the image appearance at the most local scale, whether as local gradient information, per-voxel differences, or distances and statistics on ordered collections of points. They rely on a voxel-scale correspondence between training and target and across training cases. These schemes are effective in situations where objects of interest and their surroundings have a consistent voxel-scale relationship with one another.

However, the deficiencies of the required point correspondence are evident in some medical image problem domains. Consider for example the abdomen and its organs of interest (kidney, liver). Many of the image match models presented above were designed for segmenting objects with little structural variability, such as faces and brains. The particular objects of faces and brains (eyes, nose, caudate, corpus callosum, etc.) have consistent correspondence with one another, in that their relationship is structural.

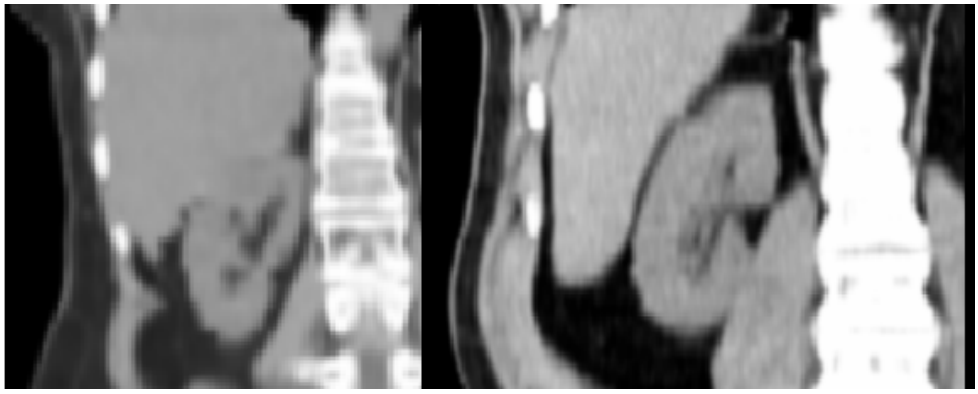

Figure 2.1: Corresponding coronal CT slices of two human abdomens. The liver, in the upper left of each patient image, has a widely disparate correspondence with the nearby kidney. This variability poses a challenge in constructing an effective image match function for prospective kidney segmentations.

However, in the abdomen there is large variability in the pose of organs relative to one another, and thus in the observed appearance, in that the organs are essentially floating (see Figure 2.1). Thus for example, an effective kidney segmentation depends on both the appearance of points interior to the kidney and also the appearance of the non-rigidly related and perhaps distorted configuration of objects near the kidney, such as liver and rib and fat.

2.2 Region-scale models

In contrast to voxel-scale methods, *region-scale* methods relax the point correspondence requirement and instead consider the image characteristics of larger scale, often anatomy-relative regions. The term image characteristic can refer to any number of properties describing a region, such as mean and variance, ranges on intensity, and the empirical distribution on intensity or image filter responses. As in the above section, the region-scale appearance models can be divided into static and probabilistic subclasses according to whether the model encodes variability in appearance. In this section I review previous region-scale appearance models and image matches.

2.2.1 Static template methods

The static template methods are those that match against a single exemplar or training case. In the image registration literature, normalized cross-correlation is a popular regional, static-template image match (Zitová and Flusser, 2003; Pratt, 1991). Recall that the goal is to determine a warp that deforms the template image to fit the target. Normalized cross-correlation (CC) is used to find corresponding (usually square) regions between the images. If X and Y are the intensities within a pair of windows, one each from the target and temple, then the normalized CC is given by

$$CC(X,Y) = \frac{E((X - E(X))(Y - E(Y)))}{\sqrt{E((X - E(X))^2)}\sqrt{E((Y - E(Y))^2)}}, \tag{2.7}$$

where $E(\cdot)$ is the expectation value or mean. CC amounts to the dot product of the windows after normalization. By finding corresponding window pairs in the images, the warp aligning the two images can be determined. While there is extensive variation in the warping model itself (Maintz and Viergever, 1998; Zitová and Flusser, 2003), the image registration literature provides a variety of image matches based on sum of squared differences, cross correlation, and mutual information.

Regional approaches most often model intensity distributions in object-relative regions, as opposed to the windows seen above. For example, the "region competition" extension to the classic active contour segmentation scheme considers the intensity distributions on the interior and exterior of the object of interest (Zhu and Yuille, 1996). In this method the local force applied at a point $\vec{v}$ on the contour is determined in part by the log-likelihood ratio of the intensities near that point,

$$\log p(I_{\vec{v}}|\alpha_{int}) - \log p(I_{\vec{v}}|\alpha_{ext}), \tag{2.8}$$

where α_{int} and α_{ext} parameterize the intensity distributions and $I_{\vec{v}}$ is the set of intensities

near $\vec{v}$. The authors model the interior with a Gaussian — $\alpha_{int} = \{\mu, \sigma^2\}$, and they model the exterior with a uniform distribution. (Yushkevich et al., 2006) claim improved results in two steps; first, rejecting the log ratio and instead using a simple difference in probabilities; and second, modeling the interior with an intensity range and two-sided smooth threshold filter and the exterior as merely the complement of the interior, $p(I_{\vec{v}}|\alpha_{ext}) = 1 - p(I_{\vec{v}}|\alpha_{int})$.

In addition to mean and variance or intensity ranges, there are many possible parameterizations of the regional intensity distribution. To this end, (Tsai et al., 2003) extend the work of (Chan and Vese, 2001) within the context of a level set-based segmentation scheme. The authors propose a gambit of regional image statistics in addition to the mean and variance, including the sum of the intensities and the sum of the squared intensities within a region. The regions in this work are the interior and exterior.

In these works, the template values are simply those image statistics for the current interior and exterior regions. The image matches or image energies described are of several types. In one example (Chan and Vese, 2001), the energy to be minimized measures the discrepancy between the inside intensities and the inside average and also the outside intensities and outside average:

$$E_{cv} = \int_{R^u} (I - \mu)^2 dA + \int_{R^v} (I - \nu)^2 dA. \tag{2.9}$$

In this example, R^u and R^v represent the interior and exterior regions and μ and ν their respective average intensities.

The template values for these image statistics can be learned in training, as they can be easily computed given expert segmentations, and the resulting image match can then be based on energies using these template values rather than the current region values. At target time the optimal segmentation is found using the gradients of these energies with respect to the shape parameters, which in this case are based on signed distance

maps.

Mean and variance and the other image statistics of (Tsai et al., 2003) are examples of *parametric* representations of the underlying regional intensity distributions. In that these low dimensional parametrics often cannot adequately represent the nuances of the observed distribution, they are simplifications that can limit the information captured by the appearance model. Some recent methods construct a static template image match using a *non-parametric* approach, modeling with a histogram the full intensity distribution for a region. These approaches require a means of comparing the template histogram with that observed at target time, analogous to the measurement of discrepancy in the image matches that use parametric statistics. Rubner et al. (2001) review and benchmark a number of these histogram matching functions with regards to classification, image retrieval, and segmentation (additional metrics can be found in (Duda et al., 2001)). Let q and r be discretized intensity distributions (histograms) between which we wish to measure the similarity, and let Q and R be the respective cumulative histograms (CDF's). Example histogram matches are found in the following table.

Name	Equation	Description
Minkowski-form	$\mathcal{L}_p(q,r) = (\sum_i \lvert q(i) - r(i)\rvert^p)^{1/p}$	L_p norm
Kolmogorov–Smirnov	$\mathcal{D}^p(Q,R) = \max_i \lvert Q^p(i) - R^p(i)\rvert$	max discrepancy on CDF
Cramer/von Mises	$\mathcal{D}^p(Q,R) = \sum_i (Q^p(i) - R^p(i))^2$	square distance on CDF
Kullback-Leibler	$KL(q,r) = \sum_i q(i) \log \frac{q(i)}{r(i)}$	summed log-likelihood

Rubner et al. (2001) also note the *Earth mover's distance* (EMD), which defines the work required to transform one histogram into another, as a powerful measure. The EMD is equivalent to the Mallow's distance metric used in the *quantile function* work of Broadhurst (Broadhurst, 2008). Quantile functions will be covered in some depth within the context of histogram matching functions in Section 2.3.1.

Leventon et al. (2000) look to capture both the interior and exterior intensity dis-

tributions, but they also consider these distributions as a function of distance from the boundary. They explicitly train the joint probability relating image intensity to signed distance from the boundary of the object of interest. They do this by using a signed distance map as their shape model and by learning the joint distribution on intensity and signed distance over a number of training cases. At target time, the image match at a place in the image is proportional to this joint probability. The algorithm locally iterates the signed distance map (prospective segmentation) so as to optimize an objective function including this image match.

The appearance models presented above, using cross correlation, image statistics, and histogram matching functions, model the variability in the intensity of a voxel within the region. However they do *not* account for the variability in the regional distribution itself. It is in that sense that the appearance models described above are termed static. Particular instances of a region, such as the interior of an organ, may have a different mean and variance, for example, whereas those methods will always search for the exemplar.

2.2.2 Probabilistic methods

There are a few region-scale image appearance models in the literature that explicitly reflect knowledge of the training population in computing a match. There are two principal approaches to this. One is to relax the static template by allowing the match to be obtained against any (or some combination of) the training exemplars. The other approach is to measure the fit of the observed regional appearance using a probabilistic distance.

In an example of the second approach, Ho (2004) uses a profile-based appearance model in the spherical harmonics segmentation framework of (Kelemen et al., 1999). The profile model is attractive because it can describe local appearance very well. However, its disadvantage is the voxel-scale correspondence it assumes, when changes in nearby

objects relative to the object of interest destroy this correspondence. The author relaxes the voxel-scale correspondence by constructing a *profile scale space*, a multiscale image match that uses blurring of the training profile data along the boundary of the object in order to disburse the local appearance around the object. At target time then, the segmentation scheme uses this multiscale model to provide a more robust localization of the object than the single scale profile model can produce.

Costa et al. (2007) propose a regional scale image match that fits against a template that changes through the course of the segmentation; this is an example of the first approach. The last level of the segmentation scheme is a deformable model stage that attempts to fine tune the mesh model result of the previous registration and morphological operations. The authors use as their initial template the histogram of the interior of the model resulting from the previous stages. The local force applied to a vertex s of the mesh is determined in part by the whether image intensities at s are compatible with the current template histogram:

$$\left| \frac{I(s) - \mu}{\sigma} \right| \leq 2 \tag{2.10}$$

This is the absolute value of the standard or "z" score. As the mesh is deformed, the histogram of the interior changes and that changed histogram becomes the template used in further deformation. This scheme uses a region-scale appearance that is probabilistic in that it is *not static*, in that the mean and variance of the particular instance is allowed to be different from what might otherwise be a static template[2].

The method of Costa et al. (2007) is like those of (Tsai et al., 2003; Chan and Vese, 2001) in that it uses low dimensional parametrics (mean and variance here) that often cannot adequately represent the observed distribution. Freedman et al. (2005) propose a "cdf distance" to match the *full* intensity distribution observed at target time against

[2]Under this understanding of a probabilistic match, the works of Chan and Vese (2001) and Tsai et al. (2003) could be included in this section.

that learned in training. The authors similarly only concern themselves with a single region, the interior of the object of interest. If q is the discretized intensity distribution (histogram) observed for a prospective segmentation and r is that learned in training, and if Q and R are the respective cumulative distribution functions (CDFs), the cdf distance is given by

$$K(q,r) = \sum_{i=1}^{n} |Q_i - R_i|^{\alpha}, \tag{2.11}$$

for some power $\alpha > 1$. As in (Tsai et al., 2003) above, the optimal segmentation is found by differentiating this image match with respect to the shape parameters, which are here the variational implicit surfaces of (Turk and O'Brien, 1999). As with (Costa et al., 2007) above, this scheme is probabilistic in the sense of the first approach. The particular training histogram r used to match against can change during the optimization to be any one of the histograms observed in training, whichever one is closest to the observed histogram q. A further complication is that the training histograms are not of the entire interior region in training, but rather random sub-regions within the interior. This approach is in the vein of nearest neighbor methods common in image classification if one thinks of the training histograms and the target histogram as existing in some space defined by the cdf distance.

2.3 Quantile functions and probabilistic region-scale appearance

In the work of (Freedman et al., 2005) and (Costa et al., 2007), there is never an explicit modeling of the variability in the regional intensity distribution. There is only an understanding that the intensity model can change to be what is currently observed in the target or to a case that was observed in training. The recent work of Broadhurst on intensity *quantile functions* (of which I have been a part) addresses this deficiency

in the appearance modeling and image match literature. The quantile function is a distribution-based regional image descriptor that has advantages over other representations, such as the histogram or CDF. The appearance models I present in Chapters 4 and 5 both depend on this representation of the regional intensity distribution and its amenability to the statistical models I use. In this section I review the quantile function construction and probabilistic image match methodology, which can be found in more detail in Broadhurst's thesis (Broadhurst, 2008) and in (Broadhurst et al., 2005, 2006).

2.3.1 Regional intensity quantile functions

Regional intensity quantile functions (RIQFs) are an image descriptor that allows one to probabilistically represent the appearance of an object in an image. Quantile functions are derived from intensity histograms within object-relative regions, such as the interior near the object boundary. The space of RIQFs has the advantage that certain common changes in a distribution, such as mean shift and intensity scaling, are represented linearly under a Euclidean metric. Principal component analysis is then used to characterize the variability in a set of quantile functions, and thus in the underlying appearance.

In the following I review the construction of the RIQF in the context of the distance metric that provides linearity. Let q and r be the continuous, one-dimensional intensity distributions in two regions between which one wishes to measure the similarity. As summarized above, (Freedman et al., 2005; Rubner et al., 2001) have proposed a number divergence measures for just this task. The Earth Mover's Distance (EMD) is one such measure that is particularly intuitive. The EMD measures the work required to transform q to r by considering the distributions as piles of dirt: the work required is the amount of dirt times the distance that it must be moved.

The Mallows distance is another such measure, proven by Levina to be equivalent to the EMD (Levina, 2002). Given q and r, with cumulative distribution functions Q and

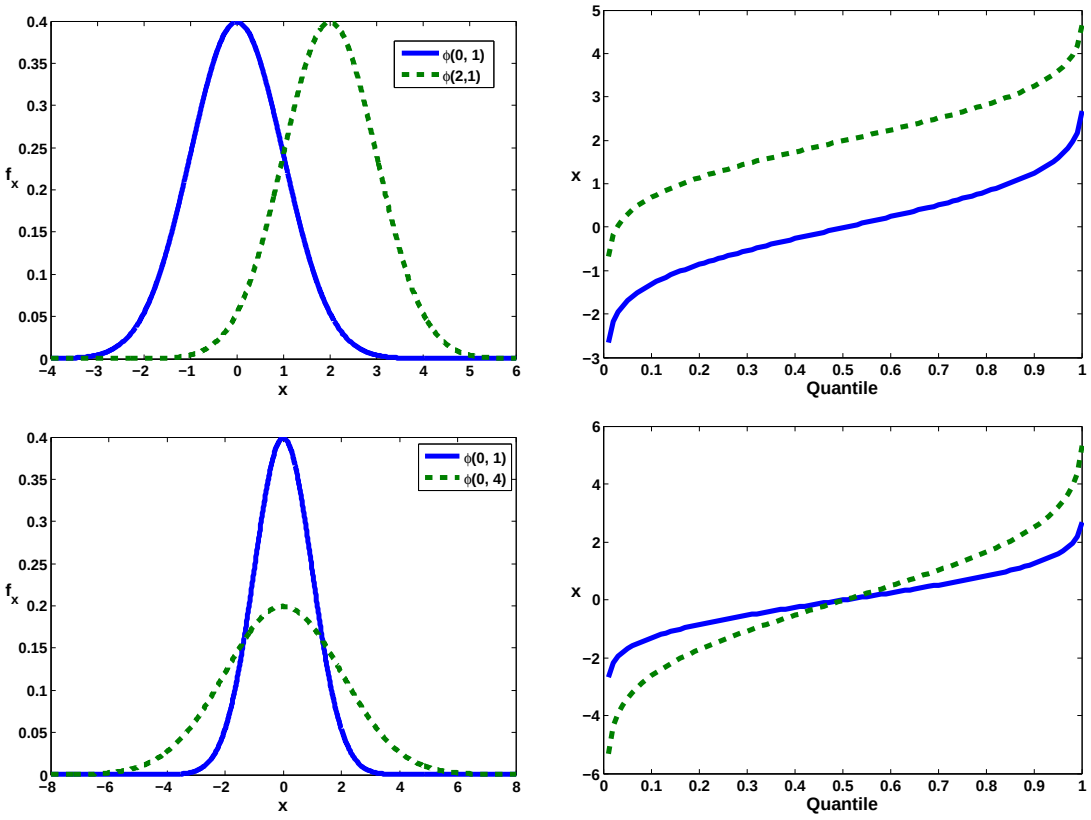

Figure 2.2: A standard normal distribution (blue), as probability density function (left column) and quantile function (right column) undergoing a mean shift and variance scaling (dashed green). Top row: mean shift. Bottom row: variance scaling. These changes are linear in the quantile feature space.

R respectively, the Mallows distance is defined as

$$M_p(q,r) = \left(\int_0^1 |Q^{-1}(t) - R^{-1}(t)|^p dt \right)^{1/p}. \tag{2.12}$$

An n-dimensional RIQF is then the discretized inverse cumulative distribution on intensities in a region, i.e., $Q^{-1}(t)$ or $R^{-1}(t)$ in the above equation. Let these discretized quantile functions be denoted $\mathbf{q}$ or $\mathbf{r}$. Coordinate j of $\mathbf{q}$ or $\mathbf{r}$ stores the average of the $[\frac{j-1}{n}, \frac{j}{n}]$ quantile of the intensity distribution for that region, i.e,

$$\mathbf{q}_j = n \int_{(j-1)/n}^{j/n} Q^{-1}(t) dt. \tag{2.13}$$

The Mallows distance above corresponds (up to a scale factor) to the L_p vector norm between $\mathbf{q}$ and $\mathbf{r}$,

$$M_p(q,r) \approx \left(\frac{1}{n} \sum_{j=1}^{n} ||\mathbf{q}_j - \mathbf{r}_j||^p \right)^{(1/p)} . \tag{2.14}$$

In the above, given a sample of data the quantile function is constructed through computing the histogram of the sample and then the corresponding inverse cumulative distribution function. I take another, equivalent approach throughout this work in my construction of the quantile function. In my work, a weight with a maximum value of 1 is associated with every sample. Using a constant weight of 1 is equivalent to having the samples unweighted. Consider a list of data/weight pairs with a total weight of N, $\{(d_i, w_i) | \sum w_i = N\}$, and a goal of constructing the associated quantile function of length n. After sorting the data such that $d_i \leq d_j, \forall i < j$, the elements of the quantile function (the quantiles) are the weighted averages of consecutive sections of the sorted list. The number of data samples in each section can be different, as long as the sum of the weights in each section is N/n. For example, the first quantile is $\frac{n}{N} \sum_i w_i d_i$, where i is such that $\sum_i w_i = \frac{N}{n}$. If N does not divide evenly, then some data samples will contribute to multiple quantiles. In the case where $N < n$, a sample could contribute to more than two consecutive quantiles.

Through the use of intensity quantile functions, distributions are understood as points in an n-dimensional Euclidean space in which distance corresponds to the M_2 metric. Furthermore, the discretized quantile function of a distribution has advantages over other representations in that some important changes are represented linearly in this space. Namely, a shift in the mean or scaling of the variance are both linear changes in the quantile function space. For example, the M_2 distance between Gaussian distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ is $\sqrt{(\mu_1 - \mu_2)^2 + (\sigma_1 - \sigma_2)^2}$ (Broadhurst et al., 2006). Figure 2.2 shows a standard normal distribution undergoing each of these changes separately.

2.3.2 Principal component analysis (PCA)

Principal component analysis (PCA) is a technique for determining modes of variation in Euclidean data and can be used to construct a probability distribution on such data. It basically finds orthogonal directions of greatest variance in the data. It is the singular value decomposition of the covariance matrix Σ, and in general solves the equation $\Sigma\vec{x} = \lambda\vec{x}$, where the satisfying λ and $\vec{x}$ are termed eigenvalues and eigenvectors of Σ respectively. Consider a sample of m n-dimensional points, $\vec{x}_i \in \Re^n$. The matrix $X = [\vec{x}_1 \dots \vec{x}_m]$ records this sample. If $\vec{\mu}$ is the row-wise mean of X (which is a column vector), then the covariance matrix is given by the outer product

$$\Sigma = \frac{1}{m}(X - \vec{\mu})(X - \vec{\mu})^T \tag{2.15}$$

If $V = [\vec{e}_1 \dots \vec{e}_n]$ is the collection of eigenvectors and $D = diag([\lambda_1 \dots \lambda_n])$ is the diagonal matrix with associated eigenvalues, then $\Sigma \cdot V = V \cdot D$. Furthermore, if the original data are Gaussian distributed, $X \sim N(\vec{\mu}, \Sigma)$, then $\vec{e}_i^T(X - \vec{\mu}) \sim N(0, \lambda_i)$ and we can model the data accordingly. That is, the mean-normalized data projected onto the i^{th} eigenvector follows a normal distribution with variance equal to the i^{th} eigenvalue. Figure 2.3 shows an example of PCA on 2D data. The resulting Gaussian model is overlaid on the data.

The above all assumed that the sample size was greater than the dimension of the data. Given $m < n$ points in $\Re^n$, PCA can recover only $m - 1$ modes. This is accomplished by setting Σ to the inner product $\frac{1}{m}(X - \vec{\mu})^T(X - \vec{\mu})$ and setting the resulting $V = (X - \vec{\mu}) \cdot V$.

Given a particular principal component decomposition, given by V and D, any example point in the space, $\vec{x}_0$, can be described using V, D, and $\vec{\mu}$. Due to the orthogonality

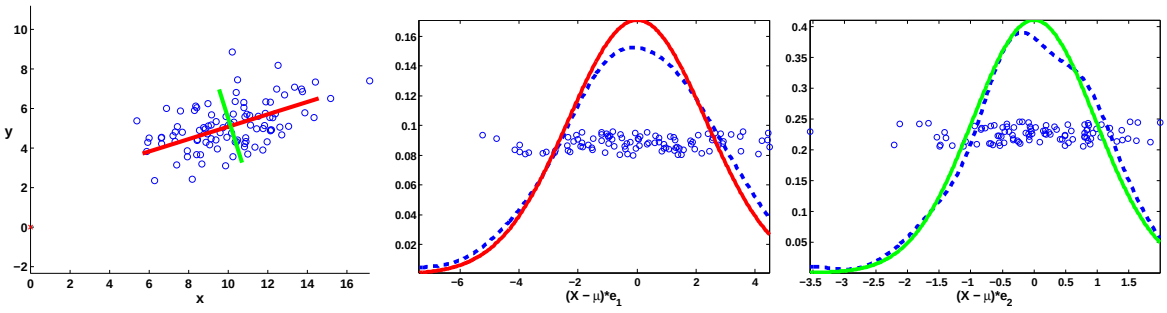

Figure 2.3: An example of principal component analysis on random 2D data. Left: scatter plot showing the data (blue) with $\pm 2\sigma$ in the principal directions plotted about $\vec{\mu}$ (red and green respectively). Middle: the data projected onto the first principal mode, $\vec{e}_1$, with a kernel density estimate (dashed blue) and the captured model $N(0, \sigma_1^2)$ (red). Right: projection onto $\vec{e}_2$.

of V, we have the equalities:

$$\vec{x}_0 = \vec{\mu} + V\vec{c}_0 \tag{2.16}$$

$$\vec{c}_0 = V^T(\vec{x}_0 - \vec{\mu}), \tag{2.17}$$

where $\vec{c}_0$ is the set of coefficients on the eigenvectors. Modeling the data as a Gaussian, we have

$$p(\vec{x}_0) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} exp\left[-\frac{1}{2}(\vec{x}_0 - \vec{\mu})^T\Sigma^{-1}(\vec{x}_0 - \vec{\mu})\right]. \tag{2.18}$$

With the data projected into the pca space via (2.17), the data covariance Σ becomes the diagonal D and the computation becomes much easier.

In that the quantile function representation allows intensity distributions to be understood as points in Euclidean space, PCA can be applied to a set of RIQFs, such as those coming from a corresponding region across training images. In this way one can construct a probability distribution on RIQFs and thus capture the variability in regional appearance. Figure 2.4 (due to Broadhurst) shows PCA on the interior and exterior intensity distributions for a bladder and prostate observed over all images of a single patient. The kind of variability captured in the observed distributions is mean

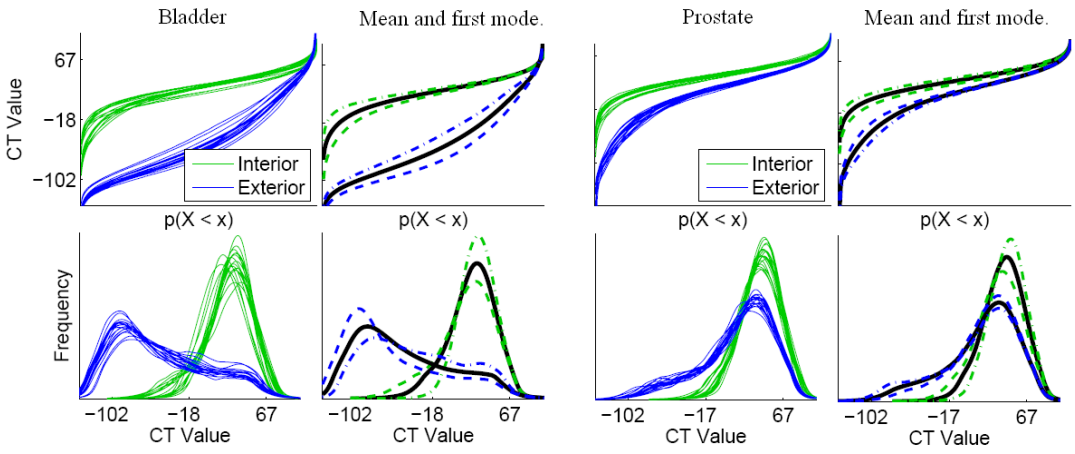

Figure 2.4: Training intensity distributions as quantile functions (top row) and histograms (bottom row). Bladder and prostate over all images of a single patient (Broadhurst et al., 2006). The regions represented are the near interior and near exterior.

shift and variance scaling, which are linear operations in the RIQF feature space.

2.4 The m-rep deformable shape model

The appearance models I develop in this dissertation are tested within the context of automatic segmentation by the posterior optimization of m-reps. This section describes the m-rep shape model and its use in Bayesian deformable model segmentation. The "m-rep", or discrete medial representation, is a skeletal geometric model used to describe the shapes of bladders, prostates, rectums, and kidneys (Joshi et al., 2001; Pizer et al., 2003, 2005a), and one can also model more complex shapes using collections of m-reps (Han et al., 2005). An m-rep is a discretely sampled grid of medial atoms, where each atom consists of a hub and two equal-length spokes. The boundary of the object model passes orthogonally through the spoke ends (see Figure 2.5).

Properties, such as spoke length and orientation, are interpolated between atoms in the grid. The model defines a coordinate system which provides an explicit correspondence between deformations of the same m-rep model (e.g., those m-reps making up a training set) and the 3D volume in the object boundary region (see Figure 2.6).

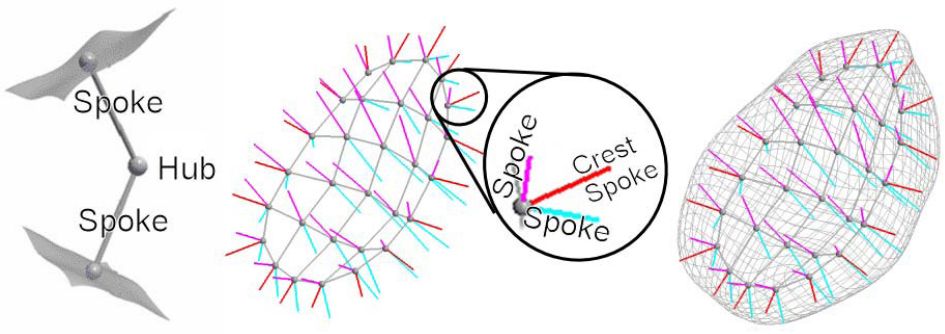

Figure 2.5: A medial atom, a discrete m-rep, and the implied boundary for a bladder.

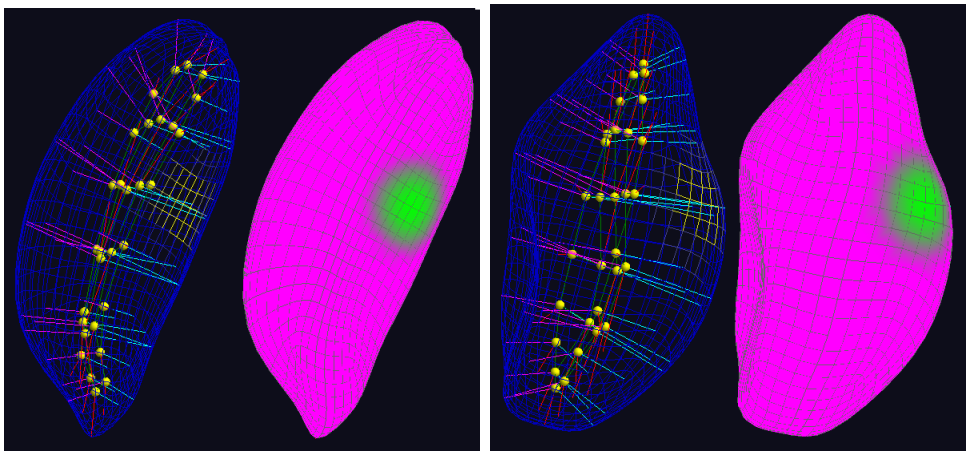

Figure 2.6: Two instances of a bladder m-rep and surface with corresponding region highlighted.

While this correspondence has proven effective when there are appropriate constraints on the deformations allowed of the model (Han et al., 2007; Merck et al., 2008), the correspondence may not be appropriate for all purposes.

The surface is a derived property of the m-rep, with the points on the surface computed deterministically from the medial grid. This leads to a constructive approach to sampling an image relative to an object, stepping along profiles normal to the surface that are also provided by the m-rep. The image regions of concern are found close to the boundary of the object of interest, as it is these regions that inform where the boundary should be.

In the Bayesian framework we segment a target image via posterior optimization of

m-reps. See Section 1.2 for a more general introduction to Bayesian deformable model segmentation. To start, a mean model (in some sense an average shape) is positioned in a target image using a similarity transform that may be computed from a few landmarks or a hand-placement or may be based on bone registration. After initialization, we optimize the posterior $p(m|I)$ of the geometric parameters given the image data. This is equivalent to minimizing the sum of the negative log-prior $\log p(m)$ and negative log-likelihood $\log p(I|m)$, which measure geometric typicality and image match, respectively. Geometric typicality and the initial mean models are based on the statistics of m-rep deformation over the training set Fletcher et al. (2004). This "principal geodesic analysis" is analogous to PCA, but on manifold rather than Euclidean data. The image match model comes from some appearance model, whether it is the template profile match described in Chapter 3 or the result of applying PCA to the set of RIQFs sampled from the m-rep training set, as is the case in the later chapters. A limited number of modes is used to avoid numerical problems and because a small number of modes is enough to describe the vast majority of variability in the data.

To compute the image match of a prospective segmentation (a given m-rep), we first sample the regions relative to the m-rep to construct the regional histogram and the resultant RIQF. The image match term is then the negative log of (2.18) after projection into the PCA space via (2.17). If $\vec{c}_0$ is the PCA coefficients of the RIQF observed for a region, then

$$-\log p(I|m) = \vec{c}_0^{\,T} D^{-1} \vec{c}_0 \tag{2.19}$$

The right-hand side $\vec{c}_0^{\,T} D^{-1} \vec{c}_0$ is seen to be the Mahalanobis distance (see Section 1.2). There is additionally a residue computed in training to model (as a 1-D Gaussian) the discrepancy between the training data and its projection into the limited PCA space. The square Mahalanobis distance of this residue is added to (2.19). The total penalty is

computed over all regions (such as interior and exterior) and the results added together for the final match.

2.5 Validation

In order to test the effectiveness my image matches, I have relied on their efficacy in segmentation tasks. I compare automatically obtained m-reps models against expert manual segmentations, which come in the form of binary images. To perform this comparison, one can either rasterize the m-rep model in order to compare directly against the binaries, or one can tessellate the boundary in the binaries and compare surface to surface. The tessellation can be accomplished using the Marching Cubes (Lorensen and Cline, 1987) technique to find the boundary isosurface in each binary image (which should be slightly blurred to smooth the voxelization). In choosing to compare surface to surface, I have developed a tool for comparing two closed surfaces called compare_byu (the surfaces are provided in the byu tile file format). This tool computes metrics such as volume overlap and closest point surface-to-surface distance in the vein of the published Valmet tool, which only compares binary images (Gerig et al., 2001). In this subsection, I will discuss validation metrics and specifically this tool and its computation of chosen metrics.

With the wealth of data being generated in segmentation studies in medical imaging, validation is an active research field (Niessen et al., 2000). Validation encompasses in part the means and measures for the quantitative and qualitative analysis of these automatically generated segmentations. My compare_byu tool provides quantitative analysis by computing multiple similarity measures between two shapes given as triangle meshes, in this case the surface from the optimized m-rep and the isosurface from the expert manual segmentation. The measures include average closest point surface distance (ASD), maximum or Hausdorff distance, and volume overlap.

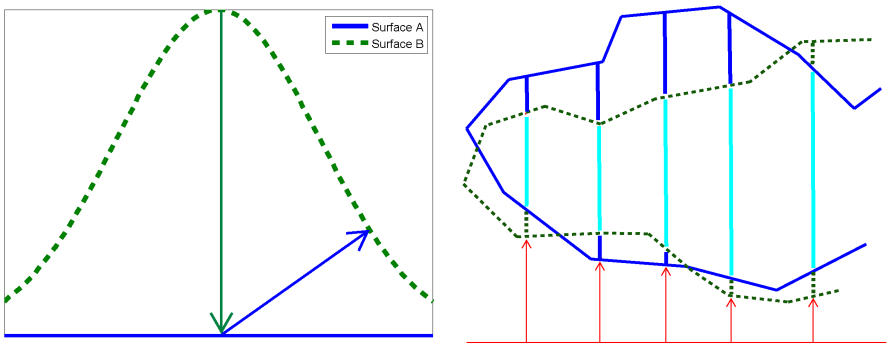

Figure 2.7: Left: The surface distance from A to B is not the same as that from B to A. Right: Schematic of the orthographic ray projections used to calculate the volume overlap. The cyan ray extents represent intersection volume, while the blue and dashed green extents represent volume exclusive to A and B respectively.

The closest point distance from a vertex a on one surface A to the other surface B is computed as the minimum over all triangles of B of the point-to-triangle distance from a (Eberly, 1999). This value is averaged over all vertices on A to obtain ASD_{AB} and its maximum is taken to obtain the Hausdorff distance from A to B. Note that $ASD_{AB} \neq ASD_{BA}$, that is ASD is not symmetric. This is seen in that the minimum distance from the vertex a to surface B is traversed along the normal of the surface B at the corresponding point b while the closest point on surface A from b will involve a normal of A, and these normals only line up if the surfaces are locally parallel (see Figure 2.7). If A is the automatically obtained surface and B is the manual reference, then ASD_{AB} may be chosen over symmetric measures (like the average of ASD_{AB} and ASD_{BA}).

Pekar et al. (2004) propose in-slice closest point distances (which are lower-bound by the 3D measure), in that in-slice distances may better represent the discrepancy perceived by the expert physicians. (Freedman et al., 2005) propose a radial distance measure, extending random rays from the center of gravity of the reference segmentation out toward the two surfaces. They report the median distance between A and B from random samplings of these rays. Costa et al. (2007) report a "robust Hausdorff" of

the 95^{th} percentile of the minimum distances observed from A to B. In that same line there may be an advantage to considering the entire distribution on minimum distances, rather than such single values as mean, median, 95^{th} and other percentiles (the median is the 50^{th}).

Volume overlap gives another performance measure, this for the degree to which the interiors of the two surfaces match up. If A and B now represent the volumes enclosed by the respective surfaces, then the most common volume overlap measure is the Dice similarity coefficient (Dice, 1945), measuring the intersection divided by the average:

$$DV(A,B) = 2\frac{||A \cap B||}{||A|| + ||B||} \tag{2.20}$$

I compute this measure between the two surfaces using a constructive solid geometry approach–a French fry method. Intuitively, consider an x-y parallel plane that is below the two surfaces, the bottom of the bounding box for example. A ray normal to the plane that extends upward will intersect the surfaces at some heights. As one moves along the ray, depending on which surfaces have been intersected and how many times, one adds bits of volume to $||A||$, $||B||$, and/or $||A \cap B||$ (see Figure 2.7). As long as the resolution of the rays on the plane is smaller than the triangles making up the surfaces, then these French fry volumes along a ray can be approximated with rectangular prisms. For speed and robustness, the actual implementation involves orthographic projection onto the three planar axes and averaging of the results.

Other accepted volume overlap measures in the literature are intersection over union, $||A \cap B||/||A \cup B||$, and intersection over reference, $||A \cap B||/||B||$ – which is also called the true positive fraction or sensitivity or recall in various clinical and other circles. The intersection over union measure is quite sensitive to the shapes being compared. Most differences generally occur at the boundary and not the interior, leading objects with a small volume relative to their surface area to have more variable volume overlaps (Gerig et al., 2001).

Additional volume measures of the the validity of a segmentation can be expressed in terms of A and B as well. For these measures, consider that ideally all the voxels of A are in B and no voxels within B are not also in A. The false negative fraction, which is the percentage of voxels in B misclassified by A, is the complement of the true positive fraction above: $1 - ||A \cap B||/||B||$. The precision or positive predictive value is the fraction of voxels in A that are correctly classified as belonging to B, $||A \cap B||/||A||$.

Chapter 3

Intensity Profile Clustering on Image Boundary Regions

In deformable model segmentation the image match can be thought of as measuring the degree to which the target image conforms to some template within the object boundary regions (see Section 1.3). The template should account for variation over a training set and yet be specific enough to drive an optimization to a desirable result. Some templates imply a constant appearance near the boundary, which is too general, while others model the appearance at each point separately, which while more specific is subject to sampling noise. In this chapter I present a novel approach to reflect the appearance changes near the boundary while lessening the sensitivity to noise of the fully local approach[1]. This method uses clustering on local intensity profiles. The method groups the appearance over many points to come up with point appearance types that robustly characterize image intensity in object boundary regions. The appearance model I present is an example of a voxel-scale, static-template model (Section 2.1.1). My method first determines local cross-boundary image profile types in the space of training data and then builds a template of optimal types. This chapter also presents results from the application of this method to kidney segmentation in CT, showing improvement over a

[1]The work presented in this chapter appeared in Stough, J., Pizer, S., Chaney, E., and Rao, M. (2004). Clustering on image boundary regions for deformable model segmentation. In *IEEE International Symposium on Biomedical Imaging (ISBI)*.

previously built model (Rao et al., 2005).

3.1 Introduction

As discussed in chapter 2, deformable model segmentations proceed based on the optimization of an objective function that includes a term measuring geometry-to-image match (image match, for short) giving the likelihood of the model with respect to the image information. The likelihood function is often trained on a set of ground truth segmentations in order to drive the optimization to expert-like results.

Among voxel-scale models (Section 2.1), several previous methods involve the use of intensity profiles as a local image descriptor. A profile is an n-tuple of intensities sampled from an image in some fixed geometry-relative way. A profile can thus be seen as a point in an n-dimensional feature space.

Previous work on building the likelihood function have included intensity profiles associated with individual image points (Cootes et al., 1995), full images referenced with a coordinate system defined by a collection of these points (Cootes et al., 1999), and intensity profiles and structure information associated with a tiled boundary (Montagnat and Delingette, 1998; Fenster and Kender, 2001). The methods typically make assumptions of a unimodal probability distribution of corresponding intensity profiles, and two of them train for each point separately. In our experience the distribution is often widely variable and multimodal, leading to concerns in the ability to model the intensity at each point by the family of training images at that point alone. To address this concern, one may choose to model the appearance at all points equally, as in looking for strong edges (Canny, 1986). This approach is oversimplified however, given the context of organ appearance in medical images.

In this chapter, I present a novel approach for training the image match, using clustering of image information in the boundary regions. The idea is that a template

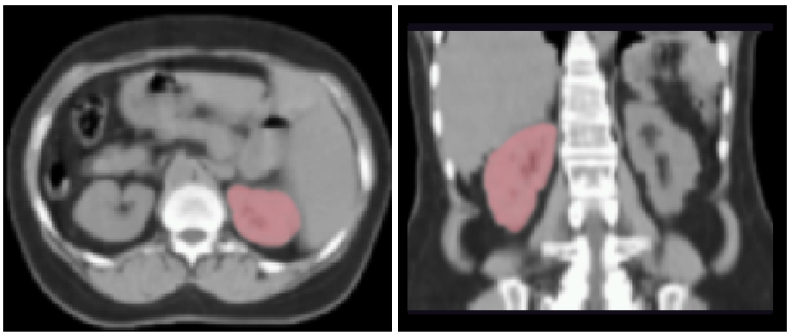

Figure 3.1: The right kidney highlighted in both axial (left) and coronal CT slices. Note the variable intensity pattern with respect to the kidney.

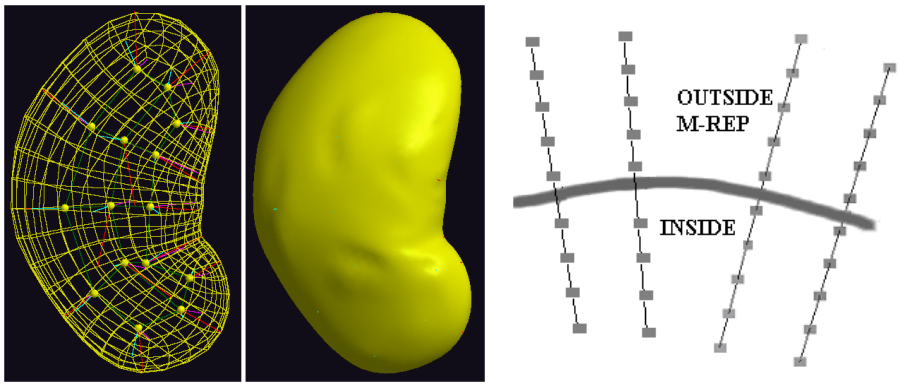

Figure 3.2: Left: the m-rep model shown as wireframe and surface. Right: diagram of closeup of the m-rep surface showing profiles, which sample intensity from inside to outside the m-rep along the normal to the boundary.

that is stable against image variability and effective in measuring image match can be made from local image descriptors that are representative of those observed in training images. As a local image descriptor, I use intensity profiles that are generated along normal directions to the densely sampled model boundary (Figure 3.2).

The shape of a particular intensity profile is dependent on both the interior and exterior of the organ of interest at the location of that profile's sampling. In that many organs have a relatively homogeneous interior, the profiles over the entire organ differ mostly in the exterior and thus are dependent on the relative position of the anatomic regions surrounding the target object. There are a limited number of such

surrounding regions, corresponding to neighboring organs, fat and bone. Figure 3.1 shows this anatomic context for the kidney. Relative to the boundary of the organ of interest, which is the center of the profile, these limited categories of surrounding region can be either abutting or nearby (distant regions are considered irrelevant). As a result, the profiles can be expected to fall into a few types depending on the profiles' sampling locations relative to these neighboring regions and their distance from the boundary. This is observed in our study of kidney CT images below (Section 3.4).

The profiles observed at a fixed object-relative point are highly variable, both as that point is sampled in different images and as small changes in the shape model affect the position and orientation of the profile sampling. The goal is to achieve a stable template appearance for image match. I consider the profiles collected over all points on the surface over a training set of images and use them to populate a high-dimensional space. I use clustering to determine profile types (cluster centers) prevalent in the data. At each boundary position on the model, a cluster center is then chosen that is the most representative of the profiles in the training set at that position. The likelihood function for new segmentations is then based on normalized correlation of the target image with a template consisting of these position dependent profile types. The template is more specific than a simple strong edge-seeking template and yet also avoids the noise problems inherent in modeling each point separately.

The method depends upon object-based associations of positions in image space provided by a geometric model. The geometric model used in this method is the medial representation, or m-rep. M-reps fitted to expert segmentations provide the spatial coordinates from which the corresponding intensity profiles are sampled in training.

Section 3.2 describes my use of the m-rep model and the image match computation. Section 3.3 describes the construction of the new image match template based on cluster centers in the profile space. An experimental study on human kidney segmentation and summary of results is presented in Section 3.4, followed by conclusions and future

directions.

3.2 M-reps and Image Match

This section provides a brief description of how I use the m-rep model, followed by details of the image match computation during posterior optimization of m-reps (refer to Section 1.2).

3.2.1 Using the m-rep model

The m-rep deformable model and its properties are reviewed in Section 2.4 and in (Pizer et al., 2005a). The clustered profile appearance model takes advantage of the local geometric coordinate system that is defined by the m-rep. First, points on the boundary of the model and their associated normal directions are given in this coordinate system. These points are the anchors for the profile sampling (see Figure 3.2). This coordinate system also defines an explicit correspondence between surface points on deformations of the same m-rep model (Figure 2.6). This dense correspondence allows us to characterize image information locally and globally.

M-reps have proven to be a powerful deformable model for segmentation via posterior optimization. In this approach the posterior of the geometric parameters given the image data is optimized. Such optimization is equivalent to optimizing the sum of the log prior, which measures the geometric typicality, and the log likelihood, which measures image match. While geometric typicality is discussed elsewhere, the image match is the topic of this chapter.

3.2.2 Image match

This section details a profile-based approximation to $\log p(I|m)$, which is the log likelihood or image match of a prospective segmentation given by m. Let I be the column

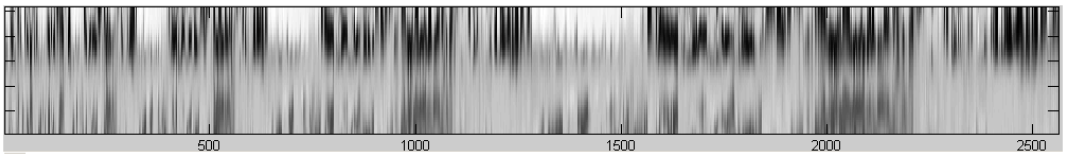

Figure 3.3: The column-wise set of intensity profiles sampled at about 2500 points on the surface of a kidney in a particular image, from inside(bottom) to outside (top) (see Figure 3.2). Note the variable profile patterns.

vector concatenation of image intensity profiles over the boundary of the model defined by parameters m, as illustrated in Figure 3.3. Assuming that $p(I|m)$ is Gaussian, the log probability is a linear function of $F(I) = -(I-\mu)^T\Sigma^{-1}(I-\mu)$, where Σ is the covariance matrix of the image data, and μ is the template image, a column vector of concatenated template profiles.

Consider the image and template, profile by profile, and make the simplifying assumption that the profiles are uncorrelated. Then Σ is block-diagonal and F(I) can be rewritten as the sum over all individual profiles I_i of the value of $-(I_i-\mu_i)^T\Sigma^{-1}(I_i-\mu_i)$ for that profile. If each profile is furthermore normalized and assumed to have uncorrelated entries, then Σ becomes the identity matrix, and $\log p(I|m)$ becomes (where I is again the collection of profiles)

$$-(I-\mu)^T(I-\mu) = -I\cdot I + 2I\cdot\mu - \mu\cdot\mu = 2I\cdot\mu - 2 \tag{3.1}$$

Thus, except for the addition and multiplication by constants, optimizing the image match under the aforementioned assumptions simplifies to maximizing the normalized correlation $I\cdot\mu$.

3.3 Building the Template

This section details the construction of the template μ from a set of training images, each equipped with a fitted boundary model, from which cross-boundary profiles can

be extracted. The goal is a template that locally represents the intensity changes near the boundary while also avoiding the noise observed locally in training. Such stable estimation usually occurs with more samples, which is limited by the size of the training set. To avoid the noise problem and limited sample size in modeling the intensity change at a point, I consider not only the profiles observed at that point, but similar profiles observed at other points. I use a clustering technique in the space of profiles to construct profile types that represent those observed over points with similar patterns. I choose representative profile types through this clustering on the collection of all training profiles. I then construct the template according to which cluster center profile is most representative of the pattern at each location. I first focus on computing the cluster centers.

3.3.1 Formation of clusters and cluster centers

The following is done in the context of boundary models with corresponding profiles and correlation-based image match. In the image match, μ is the modal image, where the match computes the target I's deviation from μ. The target is highly variable, both over different images and during the optimization itself, when a slight change of the parameters defining the surface model affects I. Thus a robust estimate of the template is in order.

First, consider the space of observed profiles, over the entire surface and over all training cases. Each profile is normalized independently. When the data is clustered, the centers of the clusters represent prevalent profile types. Normalized correlation is used as the distance metric. The clustering algorithm begins with an initial seeding of a number of profile types.

Take as an example the kidney in CT images. For initially characterizing typical image boundaries of the kidney, consider that it is a generally bright organ mostly surrounded by darker tissue (i.e, fat) but also with nearby and abutting objects that are

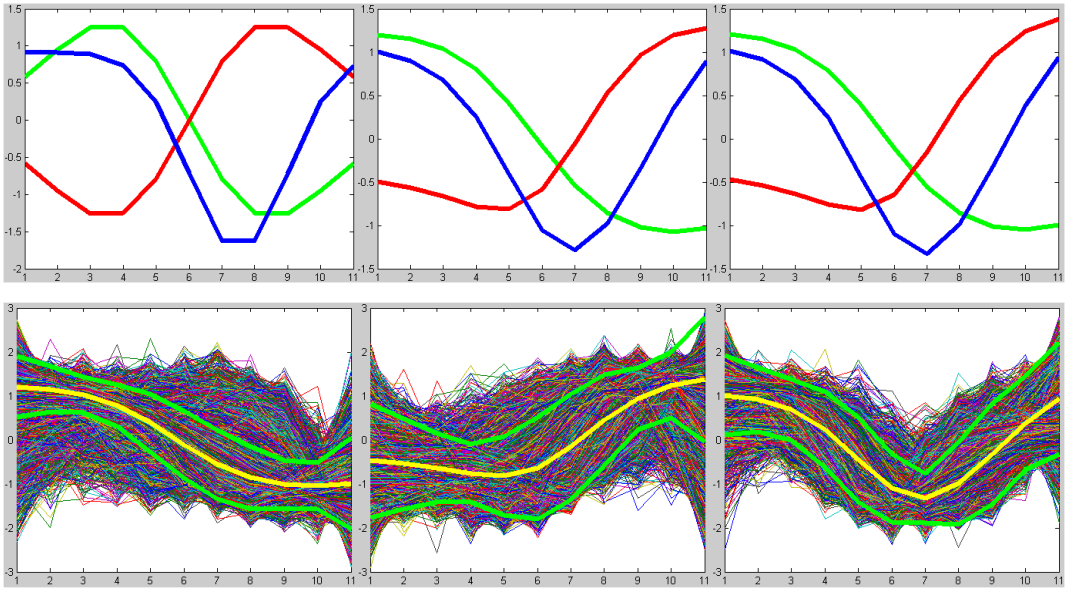

Figure 3.4: Top: the cluster centers (means) after each iteration converge to the profile types. Bottom: the training profiles grouped by cluster. The means and $\pm 2\sigma$ are highlighted.

even brighter (bone and liver or spleen). Thus one might choose the positive and negative Gaussian derivatives to characterize positive and negative edges and thus abutting darker and brighter regions respectively, and the negative Gaussian (i.e., a positive edge followed by a negative edge) to characterize the profile shapes representing nearby but not abutting bright objects (Section 3.4).

All observed profiles are binned according to their response to each type. Each profile is assigned to only one bin. The profile types are then recomputed as the mean of each bin. All profiles are compared against the new types, and the procedure iterates until convergence, when few profiles change bins between iterations. The results of the process are shown in Figure 3.4.

The resultant profile types, the bin means computed in the above algorithm's final iteration, are most representative in the least squares and correlational sense:

$$\mu = \arg\min_{\Re^m} \sum_{\nu \in \beta} ||\nu - \mu||^2 = \sum_{\nu \in \beta} (\nu - \mu) \cdot (\nu - \mu) \tag{3.2}$$

for a bin β of m-dimensional profiles. By the same simplification shown in 3.1 for normalized vectors, μ gives the highest cumulative correlation with β: the minimizer of (3.2) is the maximizer of $\sum \nu \cdot \mu$. Thus, a segmentation that maximizes the target's correlation with μ will drive to a result that is optimal over all the training data.

3.3.2 Template formation from cluster centers

In determining the local profile types, all of the profiles are tossed into a single space. However, the template used for segmentations consists of a profile at every surface position. The last step in building the template then is to use the explicit correspondence between training models to assign a profile type to each surface point. Each position is considered separately. The cumulative response of the observed profiles with each of the profile types–again, all normalized–is computed at that point over all cases. This results in a score for each profile type, the highest of which corresponds to that position's most representative profile type (see Figure 3.5). The profile type for each position is computed in the same way. The result is that the overall template is the bundle of these maximally correlated profile types. This template can then be used in future segmentations on images of the type represented in the training data.

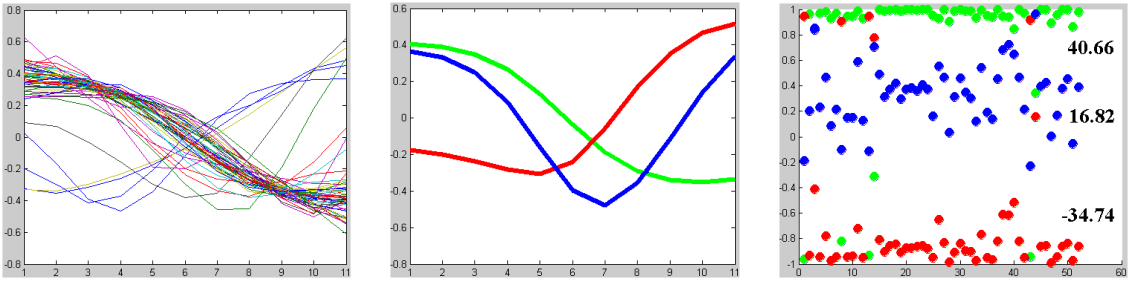

Figure 3.5: Left: profiles at a single point over all training cases. Middle: the three profile types (see Figure 3.4). Right: the training profiles' associated scores (correlations) with the profile types. Clearly, the green (light-to-dark) type yields the best response.

3.4 Example and Results

To test the efficacy of this template building approach on segmentation, Manjari Rao and I performed a study on the human kidney in CT images using m-rep deformable model segmentation. These results were compared to a previous study using a constant profile type per point, that of the Gaussian derivative. The right and left kidneys were segmented using different templates since the anatomical context is not the same on the two sides. Training of the templates was based on 52 greyscale images, with testing on 12 different CT scans. Most images had an in-plane resolution of 512×512 with voxel dimensions of 0.98 mm $\times$ 0.98 mm and an inter-slice distance between 3 and 5 mm. The images were non-contrast enhanced, acquired in Radiation Oncology at UNC on a Somatom 4+, and the test images were windowed to make the kidney intensity consistent across images. In this section I present details of this study.

Expert manual segmentations were performed on the training images using the Plan-UNC interactive contouring tool (Tewell and Adams, 2004; Chaney, 2004). This resulted in a set of binary image files, bright for kidney and dark for non-kidney voxels. An m-rep model of the kidney was then automatically deformed into each binary, leaving a set of fitted m-rep models. The template used for the image match in these binaries consisted at each point of a positive Gaussian derivative profile, with the inflection at the boundary. This is an effective edge detector for high-contrast images. By then placing the fitted m-reps into the associated greyscales, the profiles were sampled at the same model boundary coordinates for each case.

The anatomical context of the kidney organ of the abdomen is somewhat consistent across cases. For example, the right kidney usually has the liver abutting or near the superior ventral lateral corner, and a rib supports the dorsal side. The context provides a wide variation in terms of image intensity regions across the surface of the kidney. In CT scans, bone and liver and spleen (the latter superior to the left kidney) all tend to have brightness similar to the kidney, which is in turn brighter than other surrounding

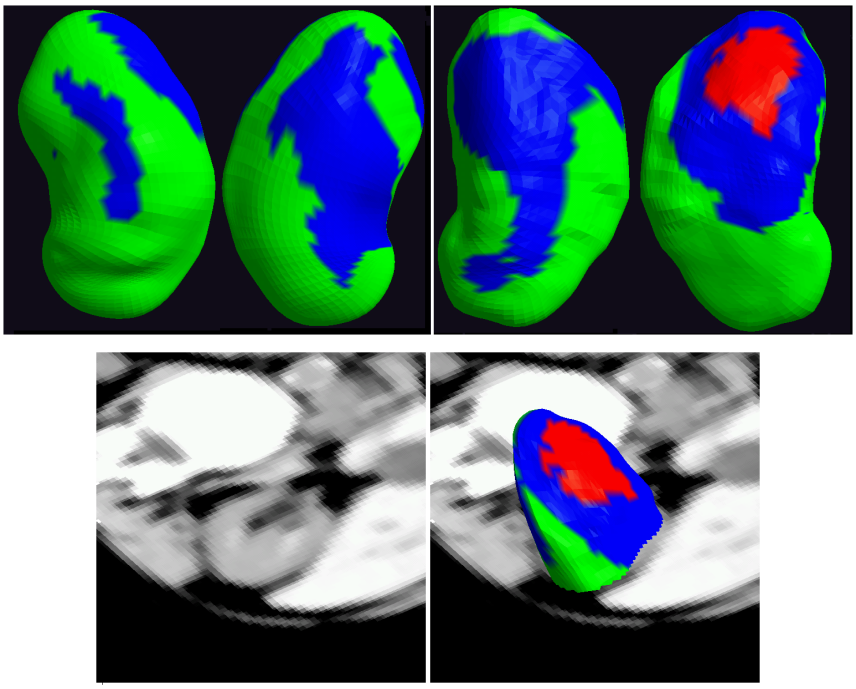

Figure 3.6: Top: kidneys colored according to the profile type selected at each point, using the same color scheme as before (see Figure 3.4). Bottom: the colored kidney model in situ in an axial CT slice. Note the grey-to-light type where the liver often abuts.

tissue. As well, in that these organs do not always abut, there can be darker regions between bright tissues. Thus as suggested in the previous section, one might expect there to be three profile types: from inside to outside the kidney, a light to dark (for most areas), a dark to light (for abutting liver and other bright tissue), and a light to dark to light again — a notch (Figure 3.1). As well, the three types can be found at similar position in the training kidneys, supporting our approach to find the best profile type to use over all cases, not for any particular case.

The three expected profile types were analytically modeled with positive and negative Gaussian derivatives and a negative Gaussian. These seeds quickly converged to prominent observed profile types (Figure 3.4). The profile type per position on the surface was then chosen based on correlation of the training data with these cluster centers (see Section 3.3.2). Figure 3.6 shows the choice per point as a coloring on the surface of

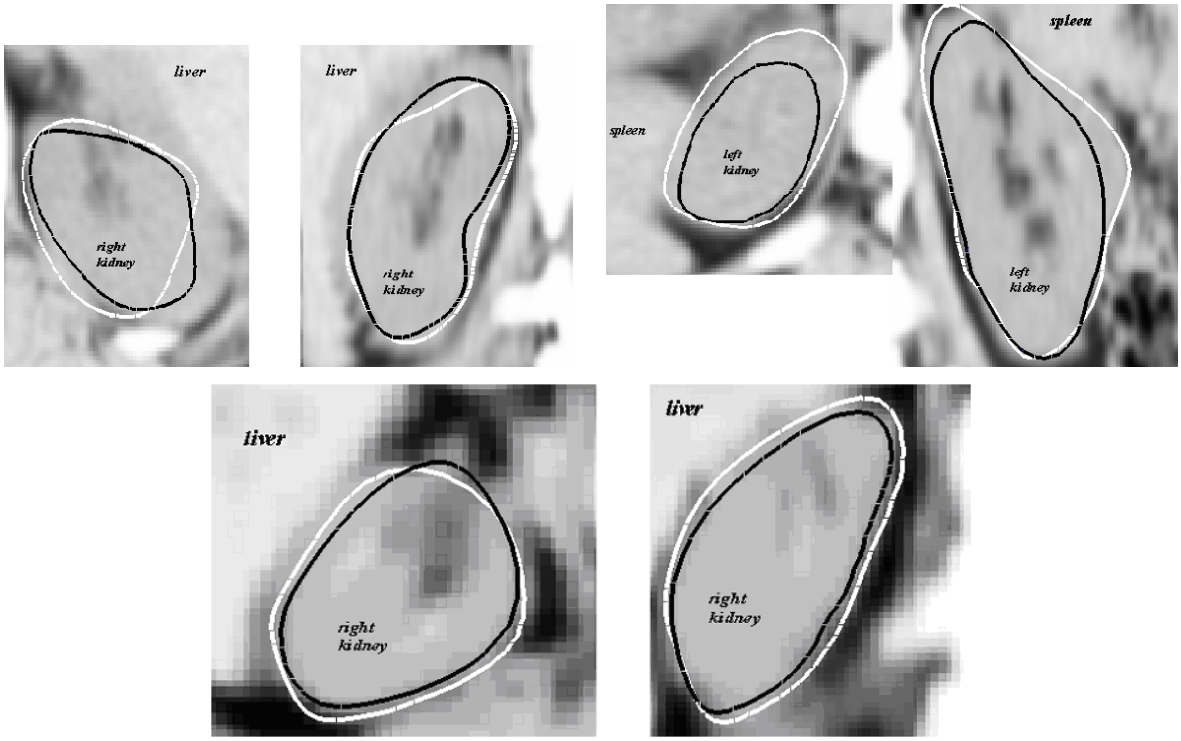

Figure 3.7: Example segmentation results on three axial/coronal slice pairs. The dark contour is the result of the Gaussian derivative template, while the lighter contour comes from the clustered template.

the kidney. The types chosen have intuitive appeal, considering general anatomy. For example, on the right kidney, the profile type chosen to represent the boundary intensity region where the liver often abuts is the dark to light type. Where bright tissue is generally nearby (within the domain of the profiles), there is the notch type.

The left and right templates were then used in m-rep automatic segmentations of the test images. For each case, the m-rep model was placed in position initially by hand, and then it was deformed automatically using the posterior optimization described in Section 2.4. The test images had two expert manual segmentations of the left and right kidneys each. To compare the m-rep model to an expert segmentation of a test case, I used the compare_byu tool described in Section 2.5.

3.4.1 Experimental results

Tables 3.1 and 3.2 summarize the median performances of the constant Gaussian derivative template (Gaus) and clustered profile template (Prof) against the two experts respectively. The comparisons are made on the basis of volume overlap (VOL), which is the intersection over union, average surface distance (ASD), and the Hausdorff distance (HAUS).

Side	GausVOL	ProfVOL	GausASD	ProfASD	GausHAUS	ProfHAUS
Left	**85.9**%	84.7	**1.25mm**	1.48	13.3	**8.30**
Right	83.3	**83.5**	**1.52**	1.70	12.4	**6.30**

Table 3.1: Kidney median overlap, surface distance and Hausdorff distance per side when comparing against expert 1. The better value of each pair is shown in bold.

Side	GausVOL	ProfVOL	GausASD	ProfASD	GausHAUS	ProfHAUS
Left	84.9%	**85.0**	**1.41mm**	1.63	11.6	**5.45**
Right	79.9	**86.0**	1.95	**1.53**	11.5	**4.95**

Table 3.2: Kidney median overlap, surface distance and Hausdorff distance per side when comparing against expert 2.

Segmentation using the cluster template approach showed an improvement in 52% of the cases when considering average surface distance to the experts and in 58% of cases considering volume overlap. The average increase in the volume overlap of automatic and expert was 1.3%. There was a marked improvement in the Hausdorff maximum distance between the two templates, with an average improvement of over 5 mm. A possible explanation for this result is that in images where the liver or spleen abuts the kidney, the Gaussian derivative template forces the geometry to the nearest strong edge, which may be far away. Figure 3.7 shows several example results obtained with the two templates.

The major improvement in using the cluster template came qualitatively, in the degree of automation. In segmenting with a template of the Gaussian derivative profile type at every position, many cases required modified parameter settings in order to

achieve the best results. Such parameters include the relative weights of geometric typicality and image match (see Section 1.2). In contrast, all parameters were constant across the segmentations when the cluster template was used. The result is that slightly improved results were obtained with significantly less human input.

3.5 Conclusions and Discussion

This chapter presented a profile clustering approach for training a robust template for image match in deformable model segmentation. I have shown the efficacy of the resulting image match in the context of automatic segmentation of the kidney in CT. The method is dependent only upon correspondence in the geometry, so the method is applicable to any model that provides it.

Simplifying assumptions were made that asserted directions for continued research. One simplification in the image match model was the choice of a single cluster center per position. As seen in Figures 3.5 and 3.8, while one profile type may be the most common at a particular position on the surface, other types may also be observed in the training cases. The scores of the observed profiles at a point with respect to all the clusters can serve as priors for a mixture model. While this is not optimal in the correlation value overall, such an approach may improve the generality of the template, leading to better results in some cases.

I also explored further qualitative improvements in the template profile image match. I normalized the target image profiles based on training data, which was found not to be preferable to the variance normalization of the profiles. I also added a per profile typical intensity penalty to complement the current match, which is based only on the shape of the profile. This provided the appearance model with a means of distinguishing sub-categories of profile types, such as those observed at positions abutting liver versus bone, where the dark-to-light pattern may itself be of a darker (liver) or brighter (bone)

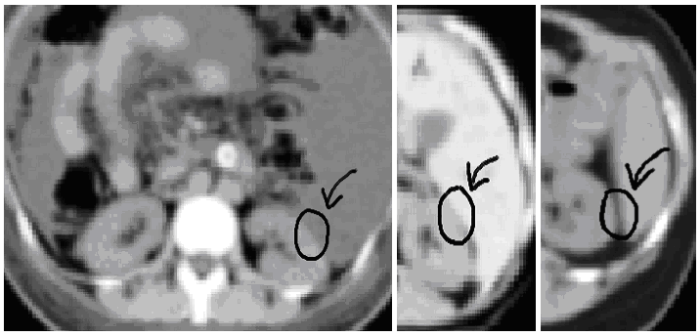

Figure 3.8: Corresponding axial slices of three CT images of the abdomen. An anatomically corresponding region may show a different intensity pattern across images. Here the kidney-liver interface is highlighted and shows all three profiles types discussed above. (Rao et al., 2005)

category.

Another problem to be addressed is the fact that the intensity profiles at neighboring positions on the model surface are almost certainly not independent as was posited (Section 3.2.2). One way to address this without hindering the simplicity of the image match computation as it stands may be to consider profiles at multiple scales. This would involve blurring along the surface to create a scale space of boundary intensity patterns, as proposed in (Ho, 2004). The correlation between neighboring profiles could then be accounted for by the information at higher scales.

A second way to address the fact that neighboring profiles are not independent is to consider an image descriptor that is more regional in nature, capturing the neighboring intensities at a point. This is the approach taken throughout the remainder of this work, where I use the intensity quantile function of Broadhurst (Section 2.3) in a similar clustering approach (see Chapter 4) and shifting regions approach (Chapter 5). In a study detailed in the next chapter, I and Broadhurst compared the profile template approach against an appearance model using near-boundary interior and exterior quantile functions. In this head-to-head test on kidney segmentation we found the quantile function-based image match led to improved results. Quantile functions also allow for a probabilistic appearance, as opposed to the static template match of this approach.

Chapter 4

Comparing Scales of Regional Appearance

4.1 Introduction

As discussed in Section 2.1.2, image match methods such as the profile approach presented in Chapter 3 rely on a voxel-scale correspondence that should not be assumed in some medical image problem domains. Furthermore, the template profile method discounts the relationship between samples at neighboring positions and the natural decomposition of the image into regions. Region-scale methods instead account for the image characteristics of larger scale regions. But what is the appropriate scale for the image match computation if voxel scale is too small?

In this chapter, I compare appearance models at three regional scales for statistically characterizing image intensity near object boundaries in the context of segmentation via deformable models[1]. The three models capture appearance in the form of Broadhurst's regional intensity quantile functions (RIQFs) (Broadhurst, 2008). These distribution-based regional image descriptors are amenable to Euclidean methods such as principal component analysis and clustering, which I use to build the statistical appearance models. To justify this regional-scale appearance, I show that an RIQF-based image match

[1]The work presented in this chapter through Section 4.5 appeared in Stough, J.V., Broadhurst, R.E., Pizer, S.M., and Chaney, E.L. (2007). Regional appearance in deformable model segmentation. In *Proc. Int. Conf. on Information Processing in Medical Imaging*, volum 4584 of *Lecture Notes in Computer Science*, pages 532-543. Springer.

leads to improved segmentations of left kidneys in CT versus those obtained with the template profile image match of Chapter 3.

The three appearance models differ in the object-relative size of the individual regions. The first model uses two regions, the interior and exterior of the organ of interest. The second model accounts for exterior inhomogeneity by clustering on object-relative local RIQFs to determine tissue-consistent regions relative to the organ boundary. The third model analyzes these local RIQFs per geometrically defined local region.

To evaluate the three models, I present segmentation results on bladders and prostates from CT in the context of day-to-day adaptive radiotherapy for the treatment of prostate cancer. Results show improved segmentations with more local regions, probably because smaller regions better represent local inhomogeneity in the intensity distribution near the organ boundary.

4.2 Motivation

Many previous image match methods depend on a voxel-scale dense correspondence across training cases. One example discussed in Section 2.1.2 is the Active Shape Model match of (Cootes et al., 1995), which considers intensity profiles associated with individual image points. The template profile match I presented in Chapter 3 also depends on a voxel-scale correspondence, where a profile type is determined at a point based on the corresponding profiles observed at that point alone in training, though the types themselves are determined through averaging over many points.

These schemes are effective in situations where objects have a consistent voxel-scale structural relationship with one another. However, in segmenting organs in the male pelvis for example, one cannot expect the same tissue-mixture at a voxel-scale correspondence (see Figure 4.1). Additionally, the profile itself is an inherently unstable image descriptor, as a 1D sampling of a part of the 3D image. A small change in the

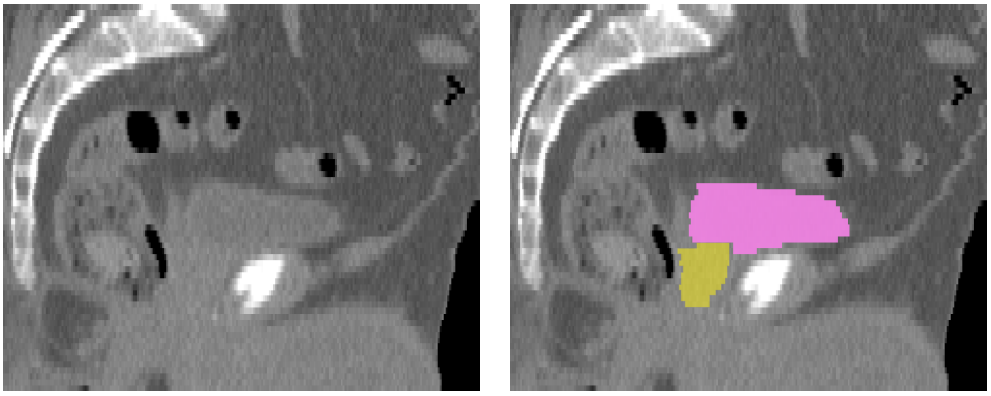

Figure 4.1: Example sagittal CT slice of the male pelvis (left), with the bladder and prostate highlighted (right). Note the lack of contrast around the organs of interest.

model parameters alters the position and normal of the profile, which can drastically change the profile observed.

Region-based methods address these concerns by modeling intensity distributions in object-relative regions. As a 3D sample of the 3D image, the regional intensity distribution provides a more stable estimate of the intensity pattern than profiles. Region-based approaches typically sample image intensities within the object interior or separately the interior and exterior. Some of the resulting intensity models use foreground/background intensity ranges (Zhu and Yuille, 1996), or use summary statistics such as mean and variance (Tsai et al., 2003; Chan and Vese, 2001).

Such simplifications of regional intensity distributions limit the information captured by the appearance model. They replace a sample of perhaps hundreds of voxel intensities with one or two numbers. Recently developed methods use an appropriate representation of the full intensity distribution for a region and compute an image match with respect to a single reference distribution (Freedman et al., 2005; Rubner et al., 2001)

The above methods do not explicitly model the variability in the regional intensity distribution. To address this deficiency, Broadhurst (2008) has developed the quantile function representation. As detailed in Section 2.3, the quantile function has the advantage over other distribution representations (such as the density or cumulative dis-

tribution function) that mean-shift and variance scaling of a distribution are represented linearly in the feature space. To the extent that these changes well describe the variation observed across training, it is appropriate to use the quantile function and, due to the linearity, Euclidean statistical methods such as principal component analysis (PCA) (Section 2.3.2) in order to build statistical models on regional intensity distributions. In an experiment described herein, Broadhurst and I showed that a quantile function based image match results in improved segmentations over those obtained using the template profile match of Chapter 3.

In (Broadhurst et al., 2006) we sampled from the global interior and exterior regions near the organ boundary and converted the resulting distributions to regional intensity quantile functions (RIQFs). The resulting image match was then based on probabilities of the regional distributions. As with previous approaches, however, the method has a significant drawback: the use of a single global exterior. Any such model oversimplifies the appearance by failing to account for the inhomogeneity in the local intensity distributions exterior to the object. While it may be reasonable to model intensities in the interior of some organs as samples from a single distribution, the exterior may consist of neighboring organs, bones, and fat and connective tissue, the intensities of which should not be considered samples from a single source. The appearance of an organ in an image is in part a function of the position and intensities of neighboring organs and volumes and their tissue mixtures.

Consider the bladder and prostate, for example. The interiors of these organs are relatively homogeneous while the exteriors consist of pubic bone, bowel, and each other. A single exterior region has the advantage of a larger sample size of voxels and thus a very stable estimation of the intensity distribution. However there is no positional sensitivity with a global region, so segmentations using a single, "global" exterior cannot account for the relative positions of these neighboring volumes, such as the bowel superior to the bladder or bone anterior to the prostate. While more local regions use fewer

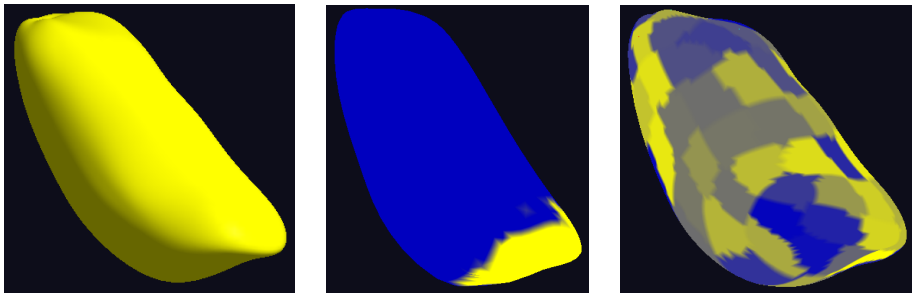

Figure 4.2: The region decomposition on the surface of the object for the three appearance models described in this chapter. Left is global, where a single region covers the entire surface. Middle is local-clustered, where local regions are combined to form larger scale regions, though not global. Right is local-geometric.

samples and thus have less stability in estimation, they provide positional sensitivity to account for these volumes. Yet as local regions get smaller, sampling noise and a lack of correspondence at fine scale begin to dominate.

In this chapter I explore this balance by experimentally comparing the efficacy in segmentation of image models at three scales. The three models use RIQFs to capture intensity distribution variability. As a baseline for comparison, the first model, global, uses two regions, the interior and exterior of the organ of interest, as in (Broadhurst et al., 2006).

The second and third models both consider the intensity distributions in a set of local regions that encode the appearance near the object boundary at fixed object-relative positions. These local regions are all of approximately the same size, with a particular region having some overlap with its neighbors, and together the local regions encompass the entire object boundary (Figure 4.2). Their projections onto the object surface are denoted "patches." I refer to the RIQF sampled from a particular local region as the associated patch's RIQF.

The second model, "local-clustered", attempts to reflect the inhomogeneity in the intensity distribution around the exterior by determining distribution-consistent region types relative to the object boundary. The region types are formed by clustering on the

RIQFs of the local regions. I then partition the object boundary according to region type and apply PCA to the local RIQFs that make up the cluster populations. For the third model, "local-geometric", each object-relative local region is treated separately. For each region, I apply PCA to the set of RIQFs sampled at that region across the training images (Broadhurst et al., 2005). Figure 4.2 shows the regional delineation of these models on the boundary of the bladder.

I also present a combined-clustered model that improves upon the local-clustered model in two ways (see Section 4.6). First, I use an improved clustering scheme that leads to cluster populations that more effectively imply the larger-scale volumes such as neighboring organs and fat. Second, I combine the local RIQFs according to the clustering to form these larger-scale regions. The combined-clustered appearance model then considers the variability in the RIQFs of the larger-scale regions rather than the variability in the local RIQFs.

Section 4.3 details the experiment comparing RIQF-based against template profile-based segmentation in a study of the left kidney in CT. Section 4.4 details the global, local-clustered, and local-geometric models. In Section 4.5 I present results of these experiments on bladders and prostates in CT in the context of adaptive radiotherapy for prostate cancer. In Section 4.6 I present the combined-clustered model along with the updated comparison on the same data set. In Section 4.7, I conclude and motivate the developments described in Chapter 5. These developments involve accounting for the variability in the conformation of the larger-scale regions implied by the RIQF cluster populations.

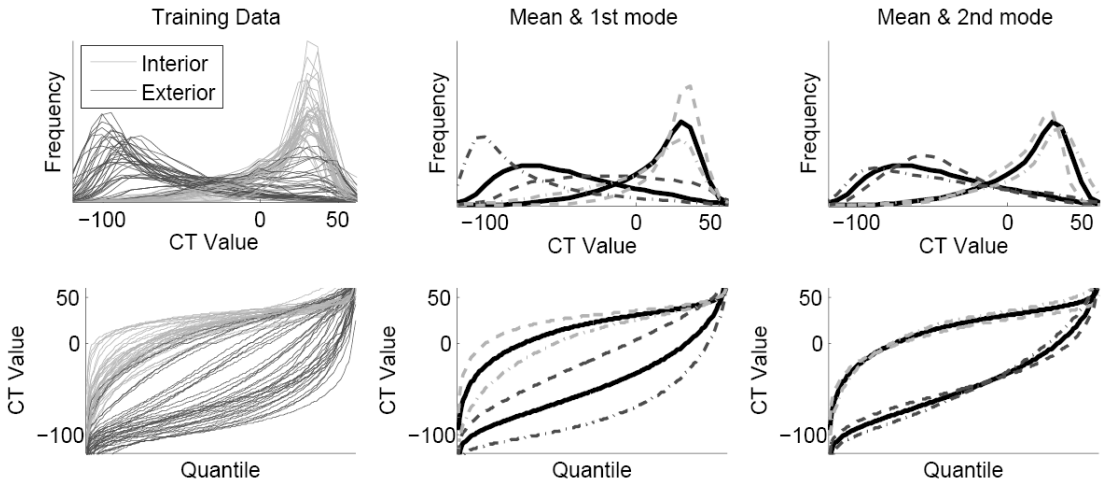

Figure 4.3: Kidney RIQF training. Top row: smoothed histograms with variation. Bottom row: the associated quantile function space, where the statics are computed (Broadhurst et al., 2006).

4.3 RIQF-based Image Match is better than Template Profile

This section details an experiment conducted by Broadhurst and me which compared a probabilistic model using global interior and exterior regions and RIQFs to the template profile model of Chapter 3. In considering the automatically obtained segmentations using the two models, this experiment shows that the RIQF-based image match leads to improved segmentations.

The data set for the experiment consists of 39 non-contrast-enhanced abdominal CT scans of different patients, each containing the left kidney, which is the organ of interest in this study. Each image has an in-plane resolution of 512×512 with voxel dimensions of 0.98 mm $\times$ 0.98 mm. The inter-slice distance of each image is between 3 and 5 mm. The automatic segmentation results were generated using a leave-one-out strategy, where training was based on all images except the target case.

In training, the left kidney of each image was segmented by an expert clinician using PlanUNC (Chaney, 2004). Six landmarks were declared on the surface of each kidney,

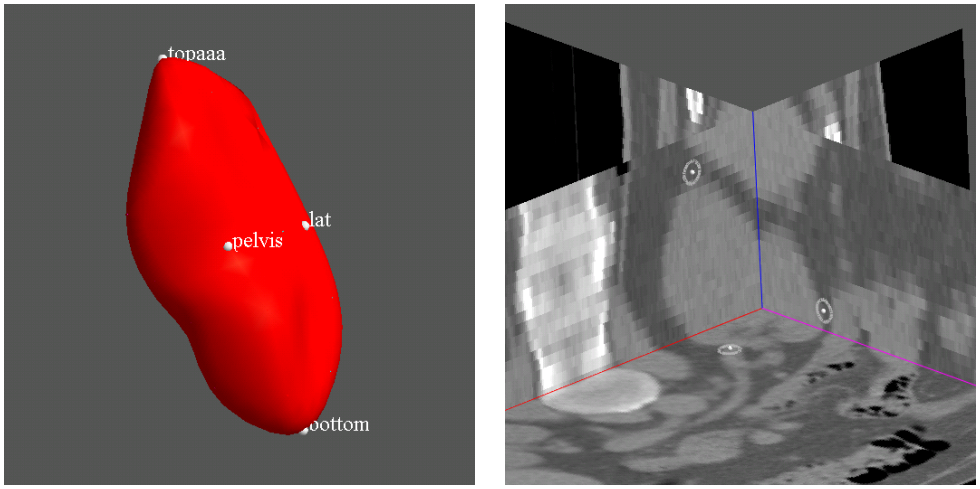

Figure 4.4: Left: Kidney m-rep model shown as sold, with four of six landmarks (two landmarks on the middle of the m-rep are obscured. Right: Orthogonal image slices showing three of the kidney landmarks.

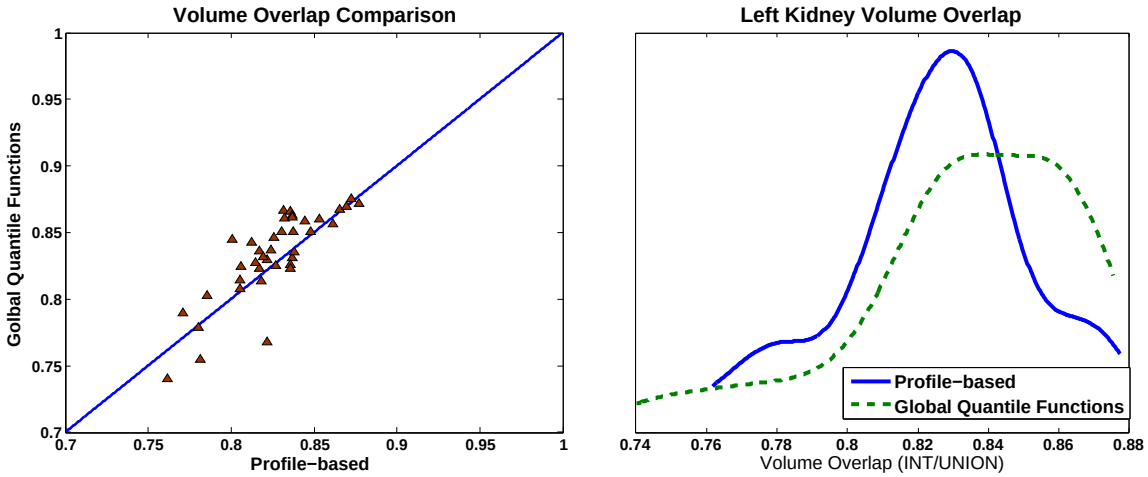

Figure 4.5: Results of the segmentation comparison. Left: Scatter plot of volume overlap of template profile model versus global RIQF model. Right: Smoothed histograms of volume overlap for the two models.

two at the superior and inferior ends of the object, and four around the middle (Figure 4.4). These landmarks and the corresponding ones on the images served to improve the correspondence of the m-reps fit to the segmentations, by forcing particular m-rep atoms to be near the image landmarks. The landmarks were also used to compute an initialization of the m-rep at target time. This is in contrast to the experiment described

in Chapter 3, where the m-reps were initialized by hand.

Considering near-boundary exterior and interior separately, the global RIQF appearance model led to the training distributions shown in Figure 4.3. In that the images used in this experiment are a subset of those used in Chapter 3, the trained template profiles were very similar to those described in Section 3.3.2.

Comparisons of the resulting segmentations are shown in Figure 4.5. Results generated using the RIQF-based match were improved over the template profile results in a majority of cases. Given the above proposed advantages and the results presented here, all forthcoming image matches discussed in this chapter and in Chapter 5 use the quantile function image descriptor.

4.4 Appearance at Three Regional Scales

Given the advantages of the quantile function representation and regional appearance modeling over voxel-scale methods, the next question is what is the appropriate scale for the image match computation. In the following sections I describe the construction of the RIQF training populations and the image match used in segmentation, for each of the global, local-clustered, and local-geometric appearance models proposed. The image match itself in each case is a sum over regions of log probabilities in the context of Bayesian deformable model segmentation (Section 2.4). The data consist of CT images of the male pelvic region and corresponding manual segmentations of bladders and prostates used for training both the shape and appearance models (Section 4.5). Each object is trained separately.

4.4.1 Global regions

With the global regions model I analyze the intensity patterns near the organ boundary, interior and exterior to it. For each training image I_p, I construct the RIQFs $\mathbf{q}_p^{in}$ and $\mathbf{q}_p^{out}$

through sampling of the image relative to an m-rep fit to the manual segmentation. The sampling approach is constructive, stepping along profiles normal to the surface that are provided by the m-rep. As a voxel is intersected by a profile, that voxel's intensity is included in a list of intensity/contribution pairs, with its associated contribution Gaussian weighted by its distance to the surface. The σ for this voxel-weighting Gaussian is a parameter of the training: this experiment used 0.4 cm, leading to a contribution of less than 0.2 (out of 1) when the voxel is farther than 1 cm from the boundary. The construction of the RIQF from a list of intensity/weight pairs is described in Section 2.3.1.

I then apply PCA separately to the two RIQF sets $\{\mathbf{q}_p^{in}, \forall p\}$ and $\{\mathbf{q}_p^{out}, \forall p\}$ to obtain Gaussian models of the intensity variation inside and outside the organ. A limited number of modes is used to avoid numerical problems, but the sum of the left-out eigenvalues is kept so as to penalize RIQFs outside of the space given by the remaining modes (Section 2.4).

For segmenting a target case, the image is similarly sampled relative to a prospective model. I treat the two regions as though they were independent, so the match computed is the sum of the squared Mahalanobis distances of the interior and exterior target RIQFs.

As per Sections 2.3.2 and 2.4 and the Equations (2.17, 2.18, 2.19), PCA on the RIQFs for each region yields a set of eigenmodes and eigenvalues, here shown as V and the diagonal matrix D. Consider the interior for example and a target intensity distribution given by the quantile function $\mathbf{r}$. Then the coefficients of $\mathbf{r}$ in the PCA space are given by $\vec{c}_{\mathbf{r}}$,

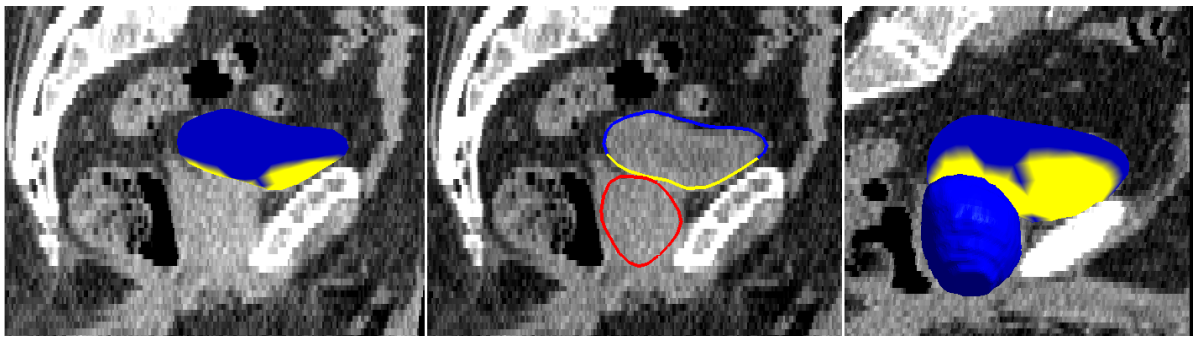

Figure 4.6: Sagittal views of a male pelvis in CT, with bladder boundary colored by region type. Left: a 3D view. Middle: the in-plane contour also colored by region type, with the prostate shown for reference. Right: an off-sagittal 3D view of the same bladder, now with the prostate also in 3D and in the foreground.

$$\mathbf{r} = \mu^{in} + V\vec{c}_{\mathbf{r}} \tag{4.1}$$

$$\vec{c}_{\mathbf{r}} = V^T(\mathbf{r} - \mu^{in}) \tag{4.2}$$

$$\hat{\mathbf{r}} = \mu^{in} + \sum_{i=1}^{n} V^i c^i_{\mathbf{r}} \tag{4.3}$$

$$-\log p(\mathbf{r}) = \sum_{i=1}^{n} (c^i_{\mathbf{r}})^2 / D_{i,i} + \frac{|\mathbf{r} - \hat{\mathbf{r}}|^2}{\sum_{i=n+1}^{N} D_{i,i}} \tag{4.4}$$

In the above, V^i is the i^{th} eigenmode. $\hat{\mathbf{r}}$ is the RIQF $\mathbf{r}$ projected into the n-dimensional PCA space. The image match penalty of the interior is set to the right-hand side of (4.4), the squared Mahalanobis distance in the first n principal directions plus the residual term. The overall penalty $-\log p(I|m)$ is the sum of the penalties over all regions, in this case the interior and exterior. The objective function being optimized consists of $-\log p(I|m)$ summed with the geometric-typicality term, also a squared Mahalanobis distance. Conjugate gradient descent optimization is used (see Section 1.2). Since $\mathbf{r}$ has been constructed through sampling relative to m, this is a match of I relative to m and not given m (Section 1.3).

4.4.2 Local-clustered regions

The impetus for the local-clustered appearance model (Stough et al., 2007) is that more local regions will better specify the exterior than the common single homogeneous region approach. The question is what constitutes a region. For my purposes, consider that an organ or other volume whose local intensity distributions are distinguishable from those of neighboring volumes constitutes such a region. Examples of such volumes are neighboring organs and fat deposits. The presence of these regions, to the extent that they are distinct from one another, is a cause of the intensity inhomogeneity in the organ exterior. I avoid modeling the 3D shapes of these regions and for my appearance model consider them only as they affect the local distributions near the boundary of the object of interest. I apply the same general notion as in Chapter 3, where I determined profile types through clustering. Here I determine region types, corresponding to these different organs and volumes, through clustering on the RIQFs of smaller regions.

Over all training images I_i, I compute RIQFs $\mathbf{q}_{ip}$ for many local exterior regions anchored to individual geometrically defined points on the object boundary indexed by p. These points are in fact the spoke ends of the m-rep model representing the object (Section 2.4). The interior of the object is treated as one region. As in the global case,

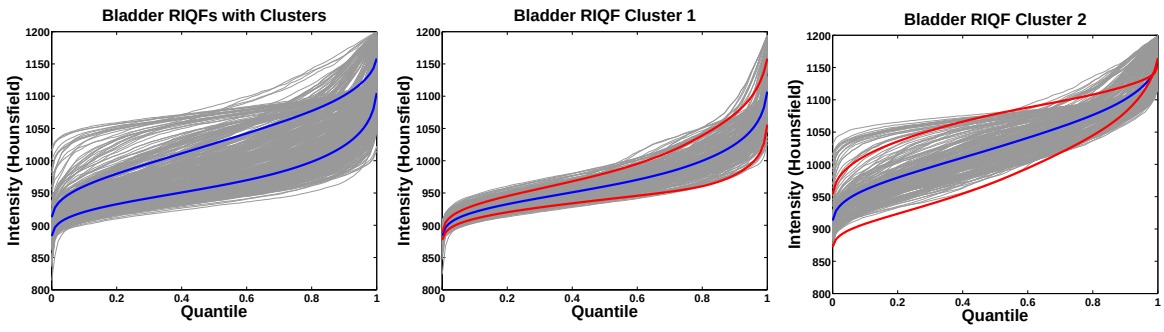

Figure 4.7: Clustering on local RIQFs for the bladder exterior. Left shows the pooled data with cluster centers overlaid (two clusters). The middle and right images show the two cluster populations with mean and ± 2 standard deviations overlaid. The reasonable separation into lighter and darker distributions is evidenced in Figure 4.6.

the contribution of a voxel to $\mathbf{q}_{ip}$ is weighted by its distance to the surface but further is equal to zero if its closest point on the surface is not close enough to point p – ensuring locality. Thus each local exterior region has its own projection onto the surface, called a patch. This patch breadth parameter is the maximum distance from the anchor point p to the projection of a voxel onto the surface for that voxel to still contribute to the RIQF of the local region anchored to p. For this experiment the patch breadth was set to 1.5 cm, which ensured that all near boundary voxels are sampled while the local regions overlap by a small amount. Parameters for this model are the density of points on the surface and the patch breadth, in addition to the distance weighting Gaussian of the global region model.

The patch breadth can also be a parameter to a continuous weighting rather than a step function. In the global model I described the distance weighting Gaussian as leading to a weight at each sample along a profile. These weights are precomputed since the profile is normal to surface and thus distance along the profile is distance from the surface. An additional patch breadth Gaussian can weight the profile itself in the following way. Each profile intersects the surface at some point, and that point has a distance to the anchor point of each local region. All of the voxels sampled along a profile can then be weighted according to which region is being sampled. The contribution of voxels sampled along this profile is then the product of the profile weight and the distance weighting Gaussian. Computing these weights involves minimal additional computation through my constructive sampling scheme.

I then cluster the pooled set of RIQFs for all boundary points and images, $\{\mathbf{q}_{ip}, \forall i, p\}$, using fuzzy c-means clustering (Bezdec, 1981). The Euclidean feature space distances inherent in this method hold for RIQFs (see Section 2.3.1). I specify the number of clusters. The results are the set of RIQF cluster centers $\{\mu^k\}$ and cluster membership scores $\{u^k_{ip}, \forall i, p\}$ over all positions and images that minimize the sum of score-weighted

distance in the M_2 metric used:

$$\{\mu^k\} = \min_{\{\tilde{\mu}_1,\ldots,\tilde{\mu}_K \in \Re^n\}} \sum_{\forall i,p} \sum_{k=1}^{K} u_{ip}^k ||\mathbf{q}_{ip} - \tilde{\mu}^k||^2, \tag{4.5}$$

with $u_{ip}^k \in [0,1]$ and $\sum_k u_{ip}^k = 1$ for given i and p. More details of the iterative procedure are presented below in Section 4.6.1.

The last two steps for training the local-clustered model are to partition the object boundary according to representative region type and to separately characterize the variability in the cluster populations. To partition the object model boundary, I use the explicit correspondence across training cases required of the shape model (see Section 4.5.1). Each point p is assigned the region type/cluster center k that maximizes $\sum_i u_{ip}^k$, the sum of cluster membership scores for that point over all images. Finally, PCA on the RIQFs making up the individual cluster populations results in Gaussian models of the RIQF variability per region type. Each position p is then assigned the PCA model of its cluster. For the interior patches, PCA is applied to the pooled set of RIQFs sampled from all points over all training images, equivalent to finding one cluster on the interior.

During target segmentation the image is sampled relative to a prospective model to populate the interior and exterior sets of RIQFs $\{\mathbf{q}_p^{int}, \forall p\}$ and $\{\mathbf{q}_p^{ext}, \forall p\}$. For each point p on the exterior, I compute the Mahalanobis distance of $\mathbf{q}_p^{ext}$ with respect to the principal modes and variances of that point's region type. For each point p on the interior, I compute the Mahalanobis distance of $\mathbf{q}_p^{int}$ with respect to the principal modes and variances found for the interior. The image match is then the sum of these terms over all points, weighted by the points' comparative importance (related to how many voxels contributed to each local RIQF).

Confirming evidence for this approach is found in the spatial distribution of region-type on the boundary that I observe in training. When I look for two clusters on the bladder data (Figure 4.7), the resultant boundary partition corresponds to lighter and

darker local distribution areas (Figure 4.6). This is anatomically justified because the bladder is surrounded mostly by lower intensity bowel and fat, with much brighter tissue from the pubic bone area and prostate inferior to it. There is similar evidence for the prostate, which has brighter tissue exterior to it in the pubic bones areas and bladder, with darker tissue elsewhere. As a result of the clustering scheme, searching for more than two clusters did not lead to such intuitive regions. Section 4.6.1 describes a new clustering on this data which does lead to cluster populations that accurately reflect the neighboring organs and volumes.

In the local-clustered scheme described in this section, through clustering on local RIQFs I determine region types which correspond to neighboring organs and volumes. The RIQFs of a cluster can be thought of as samples from one of these larger-scale volumes. The PCA model of a cluster describes the variability in these samples, which make up the set of local RIQFs belonging to that cluster. This PCA model does *not*, however, describe the variability in the RIQFs over training of the larger-scale volume associated with the cluster. In Section 4.6 below, I present an extension to the local-clustered approach that *does* account for the variability of these larger-scale RIQFs, which are constructed from the local RIQFs.

4.4.3 Local-geometric regions

A problem with the global model is that through considering only a single exterior region it sacrifices all positional sensitivity. The appearance model is not specific enough, in that there could be many large-scale regions with the same aggregate tissue mixture. The local-clustered model attempts to capture the exterior inhomogeneity and replaces a single Gaussian model of intensity with a number of local region-type Gaussian models, thus leading to a better specified appearance. However, in analyzing the RIQF data it is clear that these region-types still over-generalize the local intensity distribution at many particular points on the object boundary. Figure 4.8 shows a point's regional

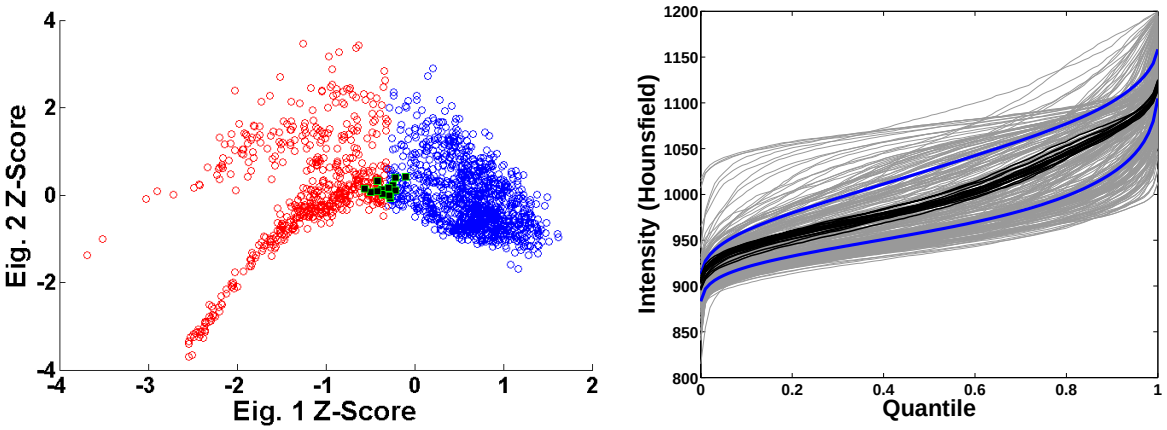

Figure 4.8: Plots showing the over-generalization of the local-clustered scheme, for bladder data. Left shows the data projected onto the first two eigen directions of the pooled PCA space, colored according to cluster type (two clusters). Right shows the pooled curve plots, with cluster centers. The highlighted data are the local RIQFs for a particular geometrically corresponding point across training. Note that either cluster population poorly represents that point's variability.

data relative to the cluster populations. Neither cluster population is representative of that point's variability; yet the RIQFs themselves are tightly grouped (at least in that projection). So one must use an even more specific model.

The local-geometric appearance model addresses the weaknesses of both previous models. In training the local RIQFs are sampled as in the local-clustered model. However, in this scheme I analyze the variability at each point on its own, applying PCA to the set $\{\mathbf{q}_{ip}, \forall i\}$ for each point p, thereby constructing a more specific Gaussian model for each local region. The interior is treated in the same way. At target time then, I treat the local regions as though they were independent. I compute the squared Mahalanobis distance for the RIQF of each point p with respect to the principal modes and variances of that point's PCA model. The image match is then the weighted sum of the squared Mahalanobis distances, as in the previous models.

4.5 Experimental Results

I compare the efficacy of the three appearance models in the context of deformable model segmentation of bladders and prostates. These organs present a very challenging segmentation problem due to the lack of contrast between the bladder and prostate and the large variability of the bladder across days (Figure 4.1). I ran the same experiment three times, with the only difference being the image match model. The data consists of five patient image sets, each of approximately 16 daily CT scans of the male pelvic area taken during radiotherapy courses. The images have an in-plane resolution of 512 $\times$ 512 with voxel dimensions of 0.98 mm $\times$ 0.98 mm and an inter-slice distance of 3 mm. I was also provided expert manual segmentations of the bladder and prostate in every image. I considered the patients separately, segmenting the images from one patient in a leave-one-day-out study, where training was based on all days for the patient except the target day. In section 4.5.1 I discuss the experimental setup. In section 4.5.2 I present segmentation results.

4.5.1 The segmentation framework

In this section I discuss the particular parameters of the experiment. I use the m-rep (see Section 2.4), which provides the means of sampling the target image relative to the prospective segmentation, as well as the explicit correspondence used in the local-clustered and local-geometric models.

To extract m-reps from images, I perform Bayesian deformable model segmentation, with a semi-automatic initialization. To start, a mean bladder or prostate model is positioned in a target image using a similarity transform computed from two prostate landmarks. After initialization, the optimization follows that described in Section 2.4 and (Fletcher et al., 2004; Pizer et al., 2003).

There are several parameters that specify the appearance models I use. The choice

Patient	GlobDSC	ClustDSC	LocalDSC	GlobASD	ClustASD	LocalASD
1	91.0%	92.0	**93.1**	1.43mm	1.50	**1.16**
2	93.5	93.6	**94.0**	1.23	1.15	**1.09**
3	90.9	91.3	**92.8**	1.58	1.48	**1.21**
4	93.7	93.9	**95.1**	1.16	1.14	**0.92**
5	89.7	89.9	**90.9**	2.13	1.98	**1.80**

Table 4.1: Bladder median overlap and surface distance per patient. See Section 4.5.2 for meaning of the abbreviations. Bold face indicates the winning method according to the measures.

Patient	GlobDSC	ClustDSC	LocalDSC	GlobASD	ClustASD	LocalASD
1	90.2%	**91.8**	90.7	0.98mm	**0.82**	0.93
2	92.0	92.3	**94.2**	1.34	1.26	**0.97**
3	92.3	92.0	**93.0**	0.95	0.94	**0.83**
4	93.9	**94.2**	**94.2**	0.97	0.93	**0.90**
5	91.3	90.0	**91.9**	1.59	1.78	**1.44**

Table 4.2: Prostate median overlap and surface distance per patient.

of region depth has already been discussed. For the local-clustered and local-geometric models I use a boundary point density that places 306 points on the bladder model surface and 290 on the prostate at fixed object-relative coordinates. I set the number of clusters to be two for both object exteriors (Section 4.4.2). My experiments show that while small changes in these parameters do affect specific results, they do not change the overall conclusions.

Gas and bone are dealt with separately from the soft tissue quantile functions. This is due to the fact that the presence of gas and bone, whose intensity distributions are very sharp and very distant from the soft tissue distributions, essentially creates mixture variability in the RIQFs (see Sections 2.3.1 and 4.6.2). To account for this, gas and bone are thresholded at intensities 800 and 1200 respectively, and their presence is accounted for in two Mahalanobis terms computed separately from the RIQF term per local region.

4.5.2 Segmentation results

I consider the relative segmentation accuracy of the three appearance models by comparing automatically generated results against the expert manual segmentations. As measures, I use average surface distance (ASD) and volume overlap given by the Dice Similarity Coefficient (DSC) (Dice, 1945), which is intersection over average (Section 2.5). I will describe the results of the global model, then the local-clustered model relative to the first, and finally the local-geometric appearance model relative to the first two. Tables 4.1 and 4.2 show bladder and prostate results for each patient and appearance model, while Figure 4.9 contains trend graphs over the pooled patient data.

The global interior/exterior appearance model results in a median volume overlap of 91.2% for bladders and 92.1% for prostates, with an overlap greater than 90% in about 50 of 80 total target bladders and 60 of 80 prostates. In terms of average surface distance, global regions results in a median ASD of 1.40 mm for bladders and 1.03 mm for prostates with 50 of 80 bladders and 65 of 80 prostates having an ASD less than 1.5 mm. These results are good in the context of the male pelvis in CT, exceeding the agreement we observe between experts.

Segmentations using the local-clustered regions appearance scheme improve upon the global results in 57.5% (46 of 80) of bladders and 53.8% (43 of 80) of prostates over all patients. Considered separately (as they are trained and segmented), this appearance model improves bladder and prostate segmentations in a majority of three of the five patient image sets. In such patient image sets, bladders are improved in 68.1% (30 of 44) of images while prostates are improved in 67.3% (33 of 49). These results are encouraging considering that even in the patient sets that are not improved in a majority of the images, the results are not significantly worse (see Tables 4.1 and 4.2).

The local-geometric model notably improves results over the first two methods, as evidenced in Figure 4.9. In every cumulative measure, this method provides better fits overall. As well, the tables show that in terms of both volume overlap and average surface

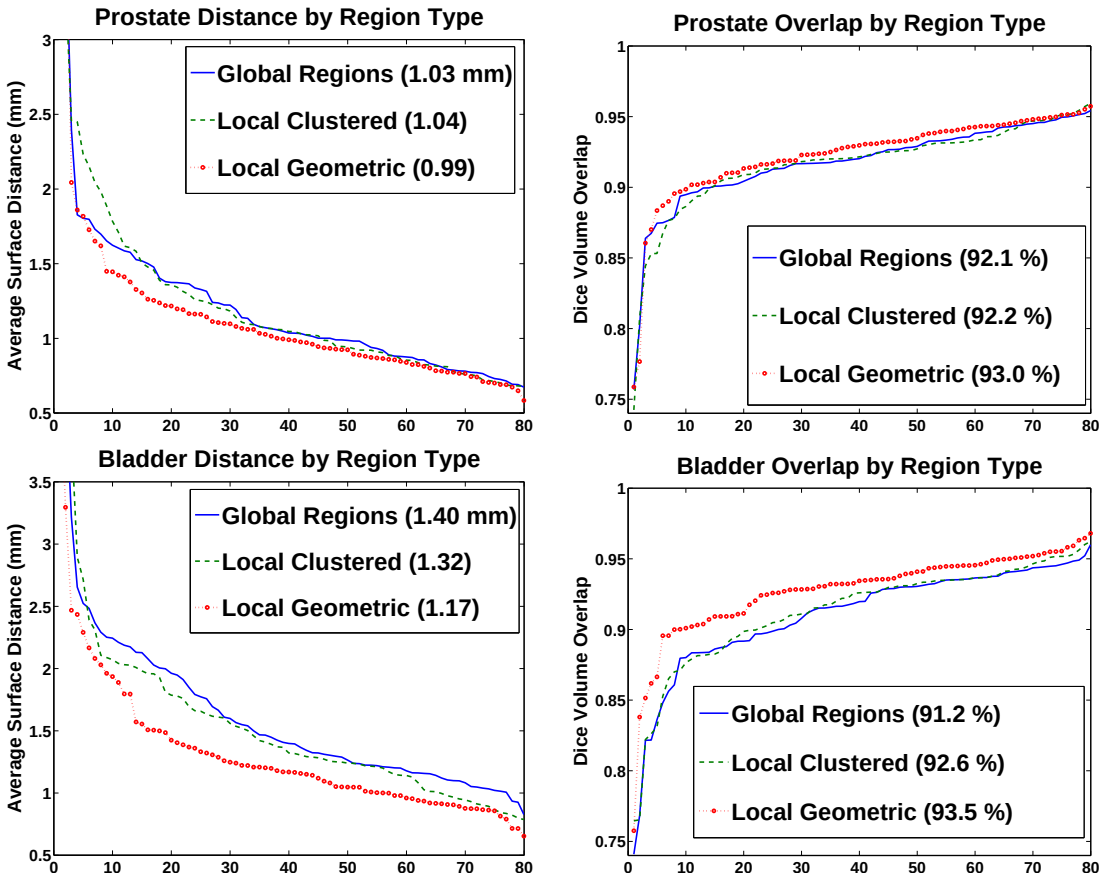

Figure 4.9: Comparisons of the three appearance models by organ (prostate and bladder by row), average surface distance (left column) and volume overlap (right column). The comparisons are with respect to expert manual segmentations. The data for each model is sorted independently over all patients to show trends, so the abscissa is image number and does not correspond across models. The numbers in parentheses are median values. While the local-clustered model just outperforms the global, the local-geometric appearance model is clearly superior.

distance measures, local-geometric is the best method in 4 of 5 patient prostate sets and in all 5 patient bladder sets. The improvements are more pronounced in the bladder because there is more room for improvement. The prostate's mostly rigid day-to-day change may be well captured by the initial transform in the segmentation algorithm (Section 4.5.1).

I also tried the local-geometric model with fewer key positions on the boundary, 78 for the bladder and 74 for the prostate, with an accompanying change in the size of each

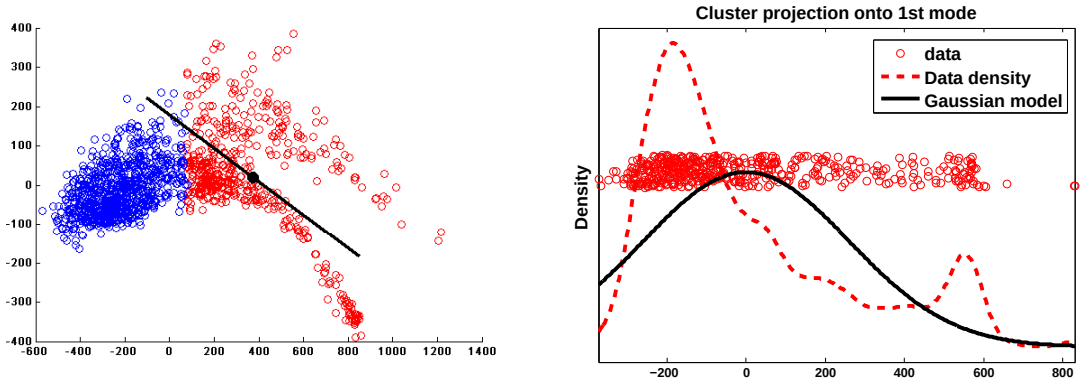

Figure 4.10: A Gaussian model applied to one of 2 clusters on bladder exterior RIQFs (see Figure 4.8). Left: Scatter plot onto the first two principal modes of the 2 cluster populations using k-means, with mean (large dot) and first eigenmode of red cluster plotted. Right: Projection of the red cluster population onto the first mode, with Gaussian model overlaid. Note that a Gaussian model for the data in this projection is not appropriate.

region to ensure full coverage. The results were slightly different from the more dense local-geometric sampling but did not change the relative performance at all. There is a benefit to this modified model, given the simplicity of the sparse sampling and that the speed of the image match computation is linear with the number of regions.

4.6 The Combined-Clustered Extension to the Local-Clustered Model

In this section I present an improved local-clustered method, called combined-clustered, that provides two advantages over the approach specified in Section 4.4.2. Namely, the new method considers the RIQF variability of the larger-scale regions implied by the clustering, and it uses an improved clustering of the patch RIQFs to determine these larger-scale regions. I will motivate these changes, present details of the improved local-clustered method, and provide updated results on the bladder and prostate. I separate these results from the above because in the intervening time many additional changes

have been made to the underlying segmentation platform.

In Section 4.4.2, I define a "region" to be a volume (such as an organ) whose local intensity distributions – given by the patch RIQFs – are distinguishable from those of neighboring volumes. However, one disadvantage of the local-clustered method is that the fuzzy c-means clustering I apply to the pooled set of local RIQFs can yield badly dispersed cluster populations. For example, the red cluster population shown in Figure 4.10 is multi-modal, leading to widely disparate RIQFs being included in the same cluster population, and thus being thought of as belonging to the same volume. In Section 4.6.1 I address this problem using the Gath-Geva clustering technique (Gath and Geva, 1989), which improves upon c-means clustering by considering the distributions of the cluster populations.

A second disadvantage of the local-clustered method is that the image match is computed as though each local patch were independent. The patches represent pieces of the larger volumes I am attempting to model through the clustering. Ideally, I should combine patches belonging to the same volume and consider the intensity distribution in that larger volume. In Section 4.6.3 I describe a means of doing this using the RIQFs for the patches. My improved clustered appearance model is then the result of PCA on the RIQFs of these larger-scale combined regions.

4.6.1 Gath-Geva clustering

In this subsection, I describe Gath-Geva clustering in the context of the RIQF training data observed at many local patches relative to an organ of interest. Gath and Geva (1989) present a classification scheme which combines the fuzzy c-means clustering of Bezdec (1981) with an additional fuzzy maximum likelihood estimation step. The scheme is actually an example of the more general "mean-shift" clustering of (Cheng, 1995), with a kernel based on the covariance estimates of the cluster populations. Upon the completion of c-means clustering, a particular RIQF must belong to the cluster

whose center is closest to it in Euclidean distance. It is the second step of Gath-Geva clustering where distance is modified according to the current covariance estimates of the cluster populations. That is, an RIQF may be closer to a particular cluster center in Euclidean distance but *less likely* with respect to that cluster population than to another.

In general, fuzzy clustering proceeds as follows. Let $\{\mathbf{q}_i, \forall i\}$ be the set of quantile functions (or feature vectors) being clustered. The initial set of centroids $\{\mu_k, k = 1 \dots K\}$ are chosen according to an either random or deterministic process. The choice of initialization can be important for the speed of convergence and for the result itself on malicious data (Su and Dy, 2004). For the data considered here, however, the resulting centroids are stable against initialization, and speed is not a concern since this is done only during training.

The degree of membership u_{ik} is computed for each vector $\mathbf{q}_i$ with respect to each cluster u_k:

$$u_{ik} = \frac{\left[\frac{1}{d^2(\mathbf{q}_i, \mu_k)}\right]^{1/(f-1)}}{\sum_{j=1}^{K}\left[\frac{1}{d^2(\mathbf{q}_i, \mu_j)}\right]^{1/(f-1)}} \tag{4.6}$$

The parameter f controls the degree to which far-away points contribute to the membership and usually is 2, as per (4.5). The denominator is a normalization so that the sum of the cluster membership scores of $\mathbf{q}_i$ is 1: $\sum_j u_{ij} = 1$. Finally, d is a distance metric and is a critical component to the final results. Initially, d is the Euclidean distance.

Given the membership scores, the centroids are recomputed using

$$\mu_k = \frac{\sum_i (u_{ik})^f \mathbf{q}_i}{\sum_j (u_{ij})^f} \tag{4.7}$$

The fuzzy clustering converges by iteratively updating the cluster centers and the set of membership scores over all the data. After this fuzzy c-means clustering initially converges, Gath-Geva clustering redefines the distance metric d to be an exponential

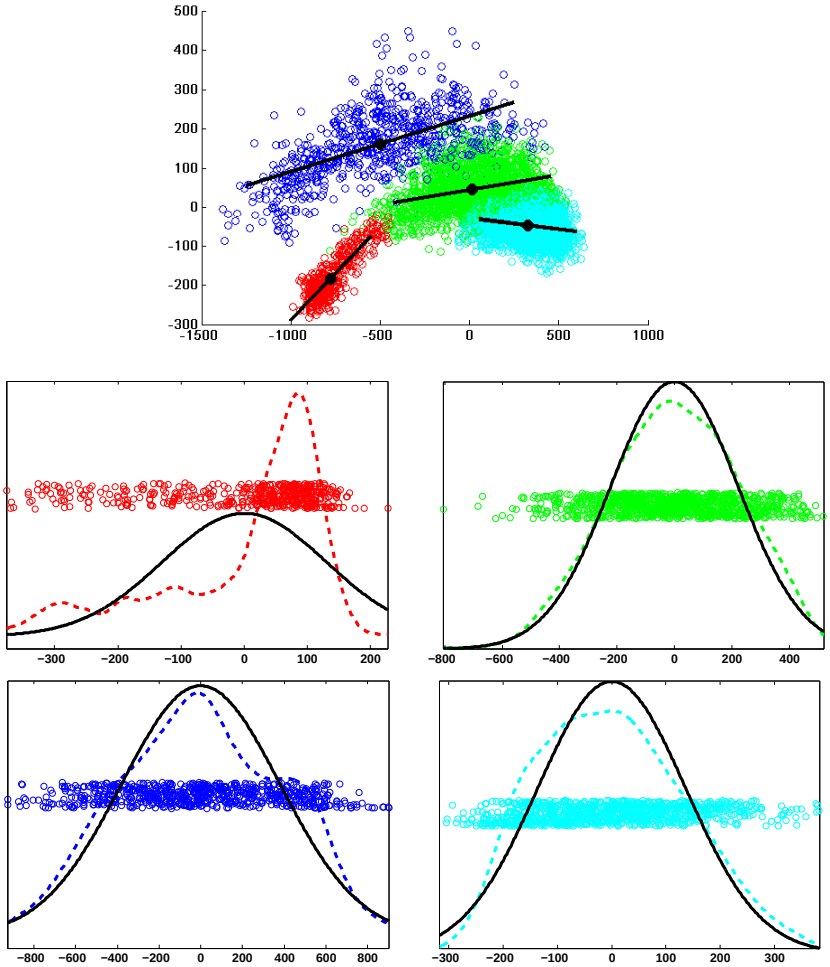

Figure 4.11: Left: Scatter plot of the 4 cluster populations using Gath-Geva, with mean (large dot) and first eigenmode plotted. Additionally, the projection data are plotted with Gaussian model overlaid. Compared to the c-means clustering, the cluster populations are more tightly distributed and better modeled with a Gaussian. A direct comparison is made in Figure 4.12.

distance computed using the covariance structure of the cluster populations:

$$h(k|\mathbf{q}_i) = \frac{1/d_e^2(\mathbf{q}_i, \mu_k)}{\sum_{j=1}^{K} 1/d_e^2(\mathbf{q}_i, \mu_j)} \tag{4.8}$$

$$d_e^2(\mathbf{q}_i, \mu_k) = \frac{[det(F_k)]^{1/2}}{P_k} exp\left[(\mathbf{q}_i - \mu_k)^T F_k^{-1}(\mathbf{q}_i - \mu_k)/2\right] \tag{4.9}$$

$$P_k = \frac{1}{N}\sum_{i=1}^{N} h(k|\mathbf{q}_i) \tag{4.10}$$

$$F_k = \frac{\sum_{i=1}^{N} h(k|\mathbf{q}_i)(\mathbf{q}_i - \mu_k)(\mathbf{q}_i - \mu_k)^T}{\sum_{i=1}^{N} h(k|\mathbf{q}_i)} \tag{4.11}$$

Here, h is the new u, the probability of selecting cluster k given the quantile function $\mathbf{q}_i$. P_k is the prior probability of cluster k and F_k is called the fuzzy covariance of cluster k, in that it is a covariance computation weighted by the posterior h.

The Gath-Geva clustering proceeds by first initializing P and F using the u determined from the fuzzy c-means clustering. Given these initializations, h is computed, followed by the cluster centers $\{\mu_k\}$, which are given via Equation 4.7 with h substituted for u. The clustering then iterates (i.e., computing P and F, followed by h and $\{\mu_k\}$) until convergence, defined to be when the maximum change in $\{h(k|\mathbf{q}_i), \forall k, i\}$ between consecutive iterations is less than some arbitrary error ϵ. For my purposes ϵ is set to 0.00001.

Given the use in the method of the covariance structure, Gath-Geva clustering leads to cluster populations that are tighter and more distinguishable from other clusters than those obtained using the fuzzy c-means. Figure 4.11 shows 4 clusters obtained using Gath-Geva on the bladder exterior RIQFs of Figure 4.8. When compared to the c-means cluster population of Figure 4.10, it is clear that the larger-scale regions are more appropriate given Gath-Geva clustering. Further evidence is shown in Figure 4.12, where the bladder surface is colored according to cluster for a particular image, for the two clustering techniques. The Gath-Geva clustering leads to regions that accurately describe the bladder neighborhood. There are cluster populations that specifically come

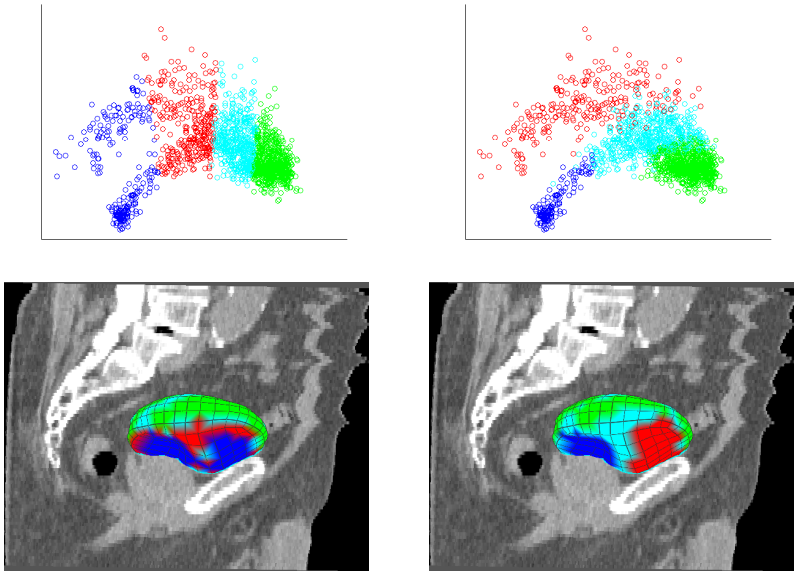

Figure 4.12: Left column: Scatter plot of the 4 cluster populations using fuzzy c-means, and visualization on the surface of cluster type per patch. Right: same, but for Gath-Geva clustering. Gath-Geva leads to more intuitive regions.

from the neighboring bone, prostate, and fat respectively. There is an additional cluster (shown in cyan) that apparently represents the mixture between the brighter prostate and bone regions and the fatty region superior to the bladder.

4.6.2 Combining patch RIQFs to form the RIQFs of larger-scale regions

In this subsection I describe the means of combining the RIQFs of local patches to determine quantile functions for larger-scale regions. The local patches that should be combined are those belonging to the same cluster, in that the cluster populations can be thought of as samples from the neighboring volumes. I want to model the intensity distribution and variability in these larger volumes, which leads to a *combined-clustered* rather than local-clustered approach.

Given two patches and their associated intensity distributions q and r, the combination of the two patches will have an intensity distribution that is simply the linear

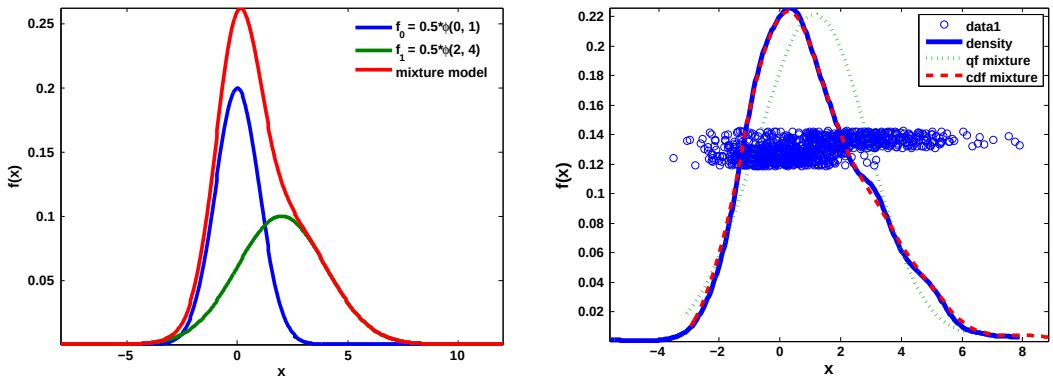

Figure 4.13: Left: Two source distributions shown in blue and green, with mixed-source distribution shown in red. Right: Random sampling from the mixed distribution shown as dots and density (in blue), with the expected density given by QF and CDF interpolation shown in green and red respectively. Clearly, CDF interpolation provides the more accurate density estimate of the combined region distribution.

combination of q and r, $(1-\alpha)q(x) + \alpha r(x)$, with the mixing parameter α denoting the fraction of voxels contributing to r relative to the total. This is due to the fact that linear interpolation in the PDF space (and CDF space) is equivalent to mixing the underlying samples (Broadhurst, 2008). However, from Section 2.3.1, mixtures are non-linear in the quantile function space. Figure 4.13 shows the QF and CDF interpolations against example mixed-source data.

If the patch RIQFs $\mathbf{q}$ and $\mathbf{r}$ and number of voxels contributing to each are the only information available, then a means for "mixing" quantile functions is called for. Broadhurst proposed embedding $\mathbf{q}$ and $\mathbf{r}$ in a higher dimensional space which includes the mixing weight α (Broadhurst, 2008). I choose instead to invert $\mathbf{q}$ and $\mathbf{r}$ to obtain CDFs Q and R, and then perform interpolation in the CDF space. I then re-invert the interpolated CDF to obtain the mixed quantile function (see Figure 4.14). I will denote this operation $\mathbf{q} \oplus \alpha\mathbf{r}$.

For the purposes of forming the larger regions from the cluster populations, only mixing with positive α is required. However the shifting regions scheme I will present in Chapter 5 will require the subtraction of a patch's QF from a larger region. For

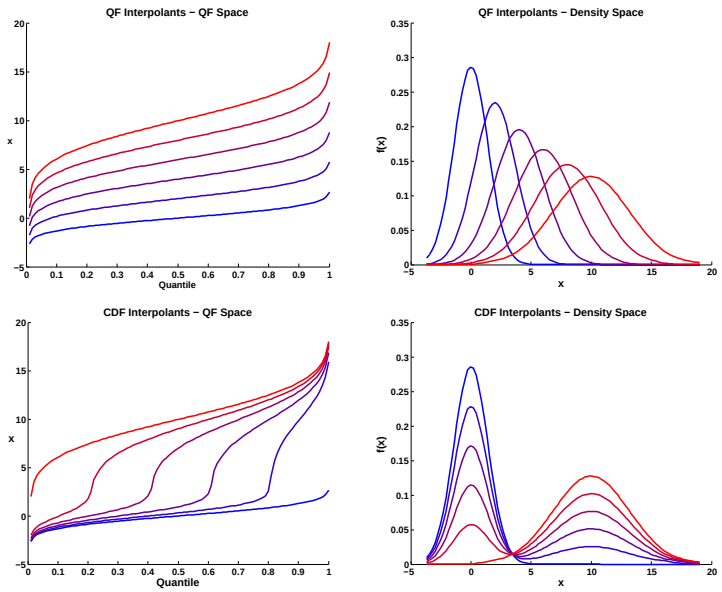

Figure 4.14: The interpolation from $N(0,1)$ to $N(10,9)$, in the quantile function space (top row) and the cumulative distribution space (bottom row). In the right column are the associated interpolants drawn as density functions. In the CDF space the interpolation leads to a mixture of the two component distributions, whereas QF interpolation leads to interpolated Gaussians. The CDF interpolation (bottom row, left) is transposed to show that it is horizontal interpolation of the quantile function plots.

convenience I will define that operation here. Consider a larger scale region with quantile function $\mathbf{q}$ and total voxel weight of N. Let a patch within the larger region be given by $\mathbf{r}$ and have a weight of $M < N$. In that it is known that the mass of r is part of the mass of q, one can subtract some fraction of r from q and still have a valid distribution (everywhere non-negative). The only issue is what is the appropriate fraction, denoted β.

If $\alpha = M/N$ then we have a $(1-\alpha, \alpha)$ mixture given by $\mathbf{q}$ and a $(0,1)$ mixture given by $\mathbf{r}$, with a goal of recovering the $(1,0)$ mixture, which I will denote $\mathbf{p}$. The quantile functions $\mathbf{p}$, $\mathbf{q}$, and $\mathbf{r}$ have representative points in the space of mixtures of $\mathbf{p}$ and $\mathbf{r}$, due to the fact that α is given and $\mathbf{q}$ is the result of mixing $\mathbf{p}$ and $\mathbf{r}$ by α. In thinking of these particular mixtures in this space, they are seen as points on the line segment from $(1,0)$ to $(0,1)$ (see Figure 4.15). Furthermore, in this space the distance between

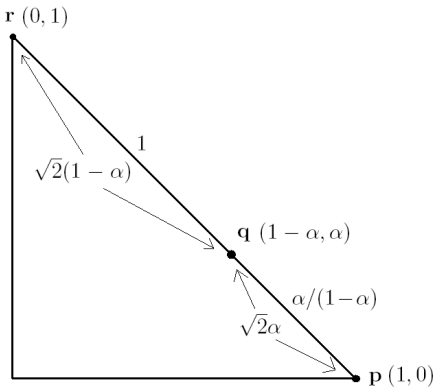

Figure 4.15: Schematic of the space of mixtures of **p** and **r**, showing the relationship between the distances in this space and the mixture coefficient α. In this space the distance between q and r is shown to be $\sqrt{2}(1-\alpha)$, and that between q and p is shown to be $\sqrt{2}\alpha$.

the mixtures represented by **q** and **r** is $\sqrt{2}(1-\alpha)$ and between the mixtures of **q** and **p** is $\sqrt{2}\alpha$. To mix the quantile functions **q** and **r** such that the result is **r** alone clearly requires a mixing weight of 1 – go all the way to **r**, or $\mathbf{r} = \mathbf{q} \oplus 1\mathbf{r}$. Thus a mixing weight of 1 is equivalent to the distance $\sqrt{2}(1-\alpha)$ on the line segment in the mixture space. Proportionately then, the distance $\sqrt{2}\alpha$ between the mixtures of **q** and **p** is equivalent to a mixing weight of $\alpha/(1-\alpha)$, but moving in the opposite direction, from **q** to **p**. It follows that in order to subtract the patch **r** from the patch **q** and thus recover **p**, one simply mixes **q** with **r** with a weight of $\beta = -\alpha/(1-\alpha)$:

$$\mathbf{p} = \mathbf{q} \oplus \beta\mathbf{r} \tag{4.12}$$

With $\alpha = M/N$, β is seen to be equal to $-M/(N-M)$.

Assumed in the above is that the patches do not share any voxels, analogous to independent sources. Since the patches in actuality have some overlap, the mixing of the local RIQFs will not give the same result as sampling the larger scale region itself, where checks are in place to ensure that any particular voxel is not counted twice. However, the method is consistent in the way it treats the larger-scale region, always as

the mixing of constituent patches.

4.6.3 Combined-clustered training and segmentation

In this subsection I describe the construction of the RIQF training population and the image match used in segmentation for the combined-clustered appearance model. The goal is a region-based appearance obtained through improved clustering of the local RIQFs as well as modeling of the intensity characteristics of the larger-scale regions given by the cluster populations.

At training time, I begin with the set of RIQFs over all images and patches $\{\mathbf{q}_{ip}, \forall i, p\}$, as per the local-clustered approach (Section 4.4.2), with the additional item noting the total weight of the voxels contributing to a patch, $\{w_{ip}\}$. I then cluster this pooled set of RIQFs using Gath-Geva clustering. The resultant $h(k|\mathbf{q}_{ip})$ term provides the class probability of each sample RIQF $\mathbf{q}_{ip}$). Each sample is assigned to the cluster maximizing h. The top image of Figure 4.16 shows the spatial distribution of this assignment for the bladder with four clusters.

At this point, the method uses the correspondence provided by the model to determine the cluster per patch. The question is which cluster is most representative of the RIQFs observed at a patch over all training images. Each patch p is assigned the cluster k that maximizes $\sum_i h(k|\mathbf{q}_{ip})$, the sum of cluster probabilities for that patch over all images, analogous to the local-clustered scheme.

Given the cluster assignment, the method now considers each training image independently and for all patches in a given image, combines those patches belonging to the same cluster. The order of the combination does not change the result, though the mixing weight α for a patch is the ratio of $\{w_{ip}\}$ to the total weight of the contributing patches so far plus $\{w_{ip}\}$, which does depend on the order. The result is the set of quantile functions for the K combined regions of each image I_i, one for each cluster, $\{\mathbf{q}_i^k, k = 1 \ldots K\}$.

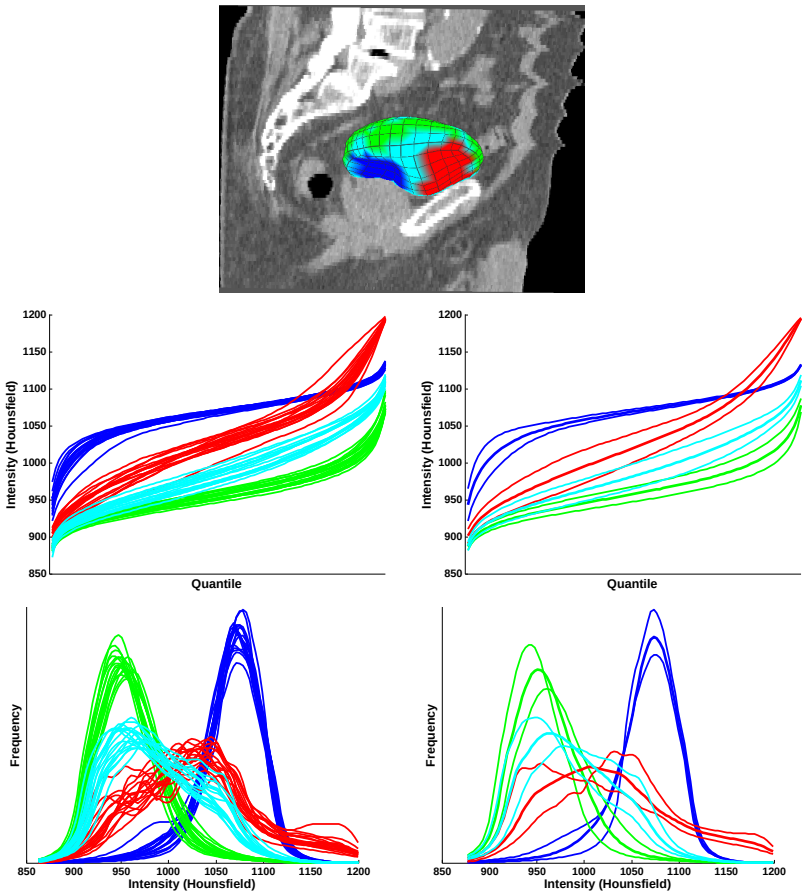

Figure 4.16: Left column: sample RIQFs for the combined larger-scale regions on the exterior of the bladder, shown in the QF and PDF spaces. Right: Mean $\pm 2\sigma$ for each of the larger-scale regions, again in the QF and PDF spaces.

The last step in training is to apply PCA to each of the K sets $\{\mathbf{q}_i^k, \forall i\}$, one per cluster k. This step results in a Gaussian model of the RIQF variability for each of the larger-scale combined regions (Figure 4.16). This is in contrast to the local-clustered method, which models the variability in the patch-scale RIQFs of each cluster.

At target time, the image is sampled relative to a prospective model to populate the set of RIQFs $\{\mathbf{q}_p, \forall p\}$. The patches are then combined according to the cluster assignments, forming the large-scale RIQFs $\mathbf{q}^k$, one for each cluster k. I then compute the Mahalanobis distance of $\mathbf{q}^k$ with respect to the PCA of that larger-scale region. The image match is then the sum of these terms over the K regions.

Patient	GlobDSC	CmClstDSC	LocalDSC	GlobASD	CmClstASD	LocalASD
1	92.6%	92.9	**93.6**	1.20mm	1.19	**1.03**
2	93.9	94.0	**94.6**	1.07	1.01	**0.95**
3	93.1	93.1	**93.4**	1.18	1.15	**1.03**
4	94.1	93.9	**94.6**	1.10	1.04	**0.98**
5	**93.1**	93.0	**93.1**	1.30	**1.19**	1.24

Table 4.3: Bladder median overlap and surface distance per patient, comparing the global, combined-clustered (CmClst), and local-geometric models. Bold face indicates the winning method for the measure used.

Patient	GlobDSC	CmClstDSC	LocalDSC	GlobASD	CmClstASD	LocalASD
1	88.8%	91.5	**92.3**	1.24mm	**0.89**	0.91
2	93.7	**95.0**	94.6	0.91	**0.81**	**0.81**
3	90.2	91.5	**93.7**	1.15	1.03	**0.80**
4	94.6	94.8	**95.1**	0.87	0.85	**0.81**
5	94.0	**94.7**	94.1	1.05	**0.96**	1.06

Table 4.4: Prostate median overlap and surface distance per patient with combined-clustered.

4.6.4 Results using the combined-clustered match

I applied the combined-clustered appearance on the same set of patient images described in Section 4.5. In this subsection I detail the results of automatic segmentation using this image match on the bladder and the prostate. Given the changes in the segmentation framework, I also retested and present here the results of the global and local-geometric image matches. Results in Tables 4.3 and 4.4 show that the combined-clustered image match does not lead to significant improvement over the global or local-geometric matches and bears a similar relationship with those matches as the local-clustered did (Section 4.5.2). Also, the results overall are improved over the original experiment, given the changes in the segmentation framework.

The experimental setup was as in Section 4.5.1 except for the following. The initialization of a model in this experiment was based on alignment of the bone in the image to one of the daily images of the patient, rather than the alignment with respect to landmarks as in the previous experiment. The alignment also affects the shape statistics,

which had an effect on the results (Section 2.4). The specific parameter settings were as follows. The sampling distance was set to 1.5 cm, with a σ of 0.4 cm for the Gaussian falloff. The patch breadth determining which samples can contribute to a local region was 0.5 cm (Section 4.4.2), with a continuous Gaussian falloff of $\sigma = 0.25$ cm. Given the improved clustering, I chose 4 clusters on the bladder exterior and 2 on the prostate exterior. The interior is treated as in the local-geometric model, with an RIQF model per point (Section 4.4.3). Gas and bone prevalence is accounted for per larger region.

Results show that the local-geometric model provided the best results in nearly all patient bladders and most prostates. Considering the prostate, the combined-clustered model led to improved median overlap in two of the five patient image sets and to improved average surface distance in three of five patient image sets. Considering both bladders and prostates, the combined-clustered model almost always improved median overlap versus the global model. Both the global and local-geometric models – which are comparable across the two experiments described in this chapter – lead to much improved results over the previous experiment. This is due to the different initialization procedure and improvements made to aspects of the segmentation framework other than image match.

I have rerun these experiments a number of times with various parameter settings. I have used fewer but larger local regions (by using different subdivision levels on the surface of the m-rep) and accounted for gas and bone per point as in the local-geometric model. Neither of these changes led to any change in the overall trend.

I also made changes in training that I thought would help. In the training described above, I use the most representative cluster at a point – determined over all training cases – as a constant choice when building the combined larger-scale regions. However, each local region of each training case is associated with a cluster, so that the combination of local regions can be based on each training image separately. This leads to tighter distributions of the larger-scale regions in training while at target time the combination

is still fixed. This change in training also did not lead to significantly different results or a change in the overall trends.

4.7 Conclusions

In this chapter, I compared the efficacy in segmentation of statistical appearance models based on RIQFs, where the models differ in the scale of the regions over which the image match is computed. The first model, the global interior/exterior regions approach, is computationally inexpensive and is the most common model in the literature. However, it is not a realistic model in many segmentation domains in that it assumes the exterior of an organ has no distinguishing features which can be useful to segmentation. Different prospective segmentations may have almost identical global regions and thus the global image match cannot distinguish between them. This can be particularly true when the prospective model is far from the right answer, for example at initialization. Some small deformation may lead to a benefit in one area while globally the change in intensity distributions is negligible, masking the effect of that deformation.

The second model considered the exterior to be made up of several large-scale regions corresponding to neighboring organs and volumes, rather than a single region as in the global case. These large-scale regions were determined through clustering on the RIQFs observed at many small-scale, local regions tied to fixed object-relative positions. In the local-clustered scheme, the image match was based on the fits of the local RIQFs observed for the prospective segmentation with respect to the cluster population assigned per point. The combined-clustered model represented a more principled approach to modeling these large-scale regions, through improved clustering and through modeling the RIQFs of the large-scale regions rather than the many local RIQFs that make up such a region. Overall, the clustering-based image matches, whether used to model the local or large-scale variability, led to better segmentation results than the global scale

image match.

However, the clustering models also impose an inaccurate appearance on many particular object-relative regions by including the variability of similar RIQFs sampled from geometrically non-corresponding positions. Whereas the clustering leads to a more stable appearance for many regions with high variability, for positions whose training RIQFs have relatively low variance, the clustering leads to broadening the distribution of possible RIQFs. Yet it is these positions of low variance that a segmentation should key off of, where there is a high degree of confidence about the expected appearance.

The final scale of appearance model I presented, the local-geometric, avoids this deficiency by analyzing the RIQFs for each local region separately. This model provides the most specific local appearance at every point by using only the RIQFs coming from that point in training, and it results in the best segmentations overall.

A key to improving the performance of the regional methods, whether local or larger-scale, is correspondence. The geometric correspondence the methods rely on is a baseline for finding what local image intensity distributions to expect, but the *image* correspondence comes from how the neighboring objects are configured with respect to the object of interest. The image correspondence is in general too strongly tied to the geometric correspondence. For example, the region representing the pubic bone may change position relative to the bladder surface across days, while the geometric correspondence of the bladders will not. The result is a false association of bone type and non-bone type intensity distributions based upon the geometric correspondence that is a source of confusion in the appearance model. I address this problem in the next chapter through a shifting region approach, where I allow changes in the region type partition on the object surface according to the image data, whether at the local-geometric or combined-clustered scale.

Chapter 5

Shifting Regional Appearance

5.1 Introduction

In Chapter 4 I claimed that the large-scale external regions found near the object boundary are found at different object-relative places in different images. This is due to the fact that the objects often imaged (such as in the male pelvis in CT) do not have structural relationships that would fix their relative positions. Yet these external regions can be key to finding the boundary of the object of interest. The correspondence provided by the geometric model often is determined relative only to the organ of interest and not those neighboring organs and volumes that affect the external appearance. Thus this external variability is not accounted for by any appearance model relying solely on the geometry.

In this chapter I propose two novel image matches which both allow shifting of the image model on the boundary to account for the variable object-relative positions of these regions. This is done within the context of the quantile function image descriptor (Section 2.3) on m-rep-defined object-relative regions (Section 2.4).

The first model, shifting-combined-clustered, extends the combined-clustered model (Section 4.6) by allowing a reasonable starting regional delineation (given by the cluster chosen per position) to deform according to the possible object-relative positions of the

region over training. The second model, shifting-local-geometric, extends this idea to the local-geometric image match (Section 4.4.3). The second extension can be thought of as the first model, but with the clusters chosen to be the RIQFs from each local region rather than the result of clustering on the RIQFs over all local regions.

To evaluate these shifting image match models, I present segmentation results on bladders and prostates from CT in the context of day-to-day adaptive radiotherapy for the treatment of prostate cancer. The segmentations achieved using the shifting models are slightly improved over their static counterparts (Section 5.5) among already very good results, which suggests that the methods may be useful but that there is little room for improvement in the test set.

5.2 Motivation

In this section I motivate the use of a shifting appearance model to relax the rigid correspondence dictated by the geometry and account for the variability in the conformation of volumes external to the object of interest. Key to an effective image match is determining *what* to look for and *where* relative to the object. Ideally, there are object-relative positions (the where) whose intensity characteristics (the what) are stable across training, with low variance, and distinct from those of neighboring positions. Then at target time, the image match can simply test the degree to which the intensity at each such key position matches those intensities seen in training at the corresponding position. Depending on the scale of these positions, correspondence is critical to this process.

In Chapter 4 I showed that region-scale appearance models have an advantage over voxel-scale models in that they consider 3D regions (Section 4.2). The correspondence across images – determined by the geometric model – is more reasonable between larger-than-voxel-scale regions than between voxels. This is so because one cannot expect the

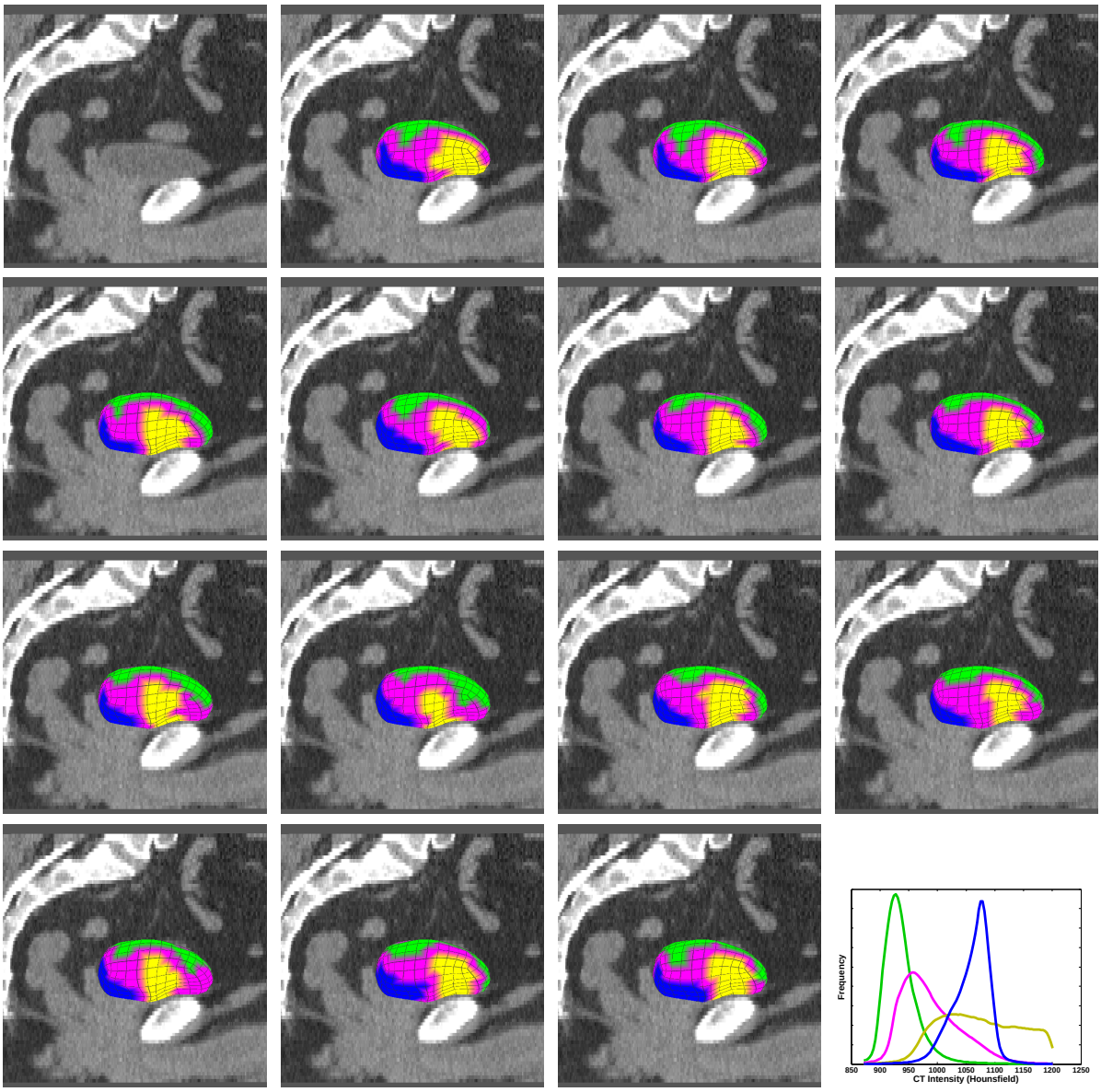

Figure 5.1: Series of sagittal images showing the changing object-relative regional conformation across images. Shown are the cluster delineations on the static surface of a bladder, for a single patient across that patient's treatment days. For context the static background image shown leads to the delineation in the second image of the first row. Bottom-right: the intensity distributions of the larger-scale regions, colored accordingly.

same intensity characteristics at a voxel-scale correspondence in the image sets of interest, male pelvis in CT. Given this confidence in the geometry-implied correspondence at a larger-than-voxel scale, the methods of Chapter 4 then consider the intensity characteristics for corresponding regions. The populations of these intensity characteristics

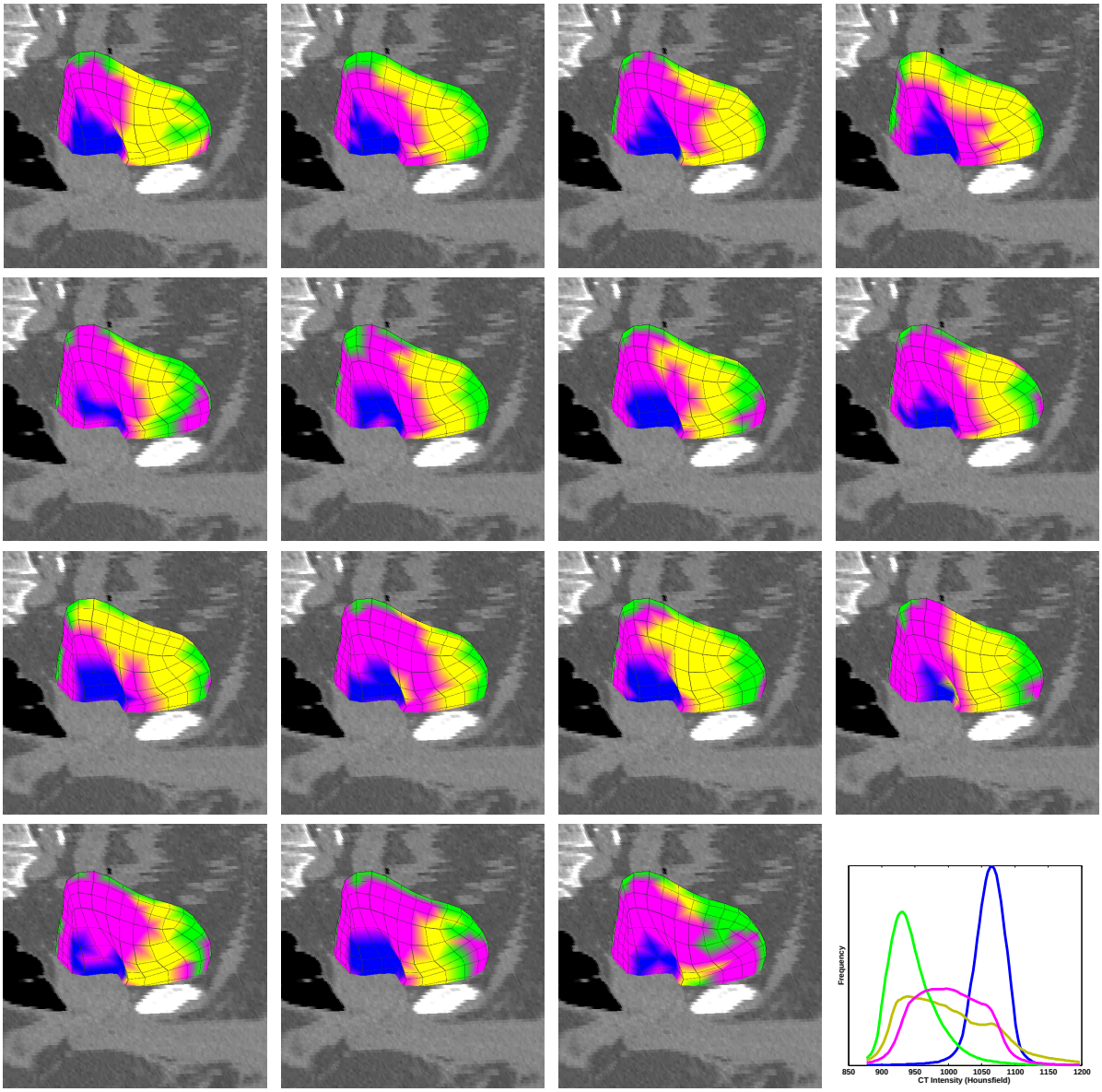

Figure 5.2: Shown from the perspective of the bottom-right of the patient, images showing the changing object-relative regional conformation across images for another patient.

are modeled via Gaussians in the space of regional intensity quantile functions (RIQFs) sampled from sets of these corresponding regions (Sections 2.3.1, 4.4).

In addition to global, local-clustered, and local-geometric appearance models, I presented and tested the combined-clustered model. This model considered a larger-scale region to be a volume (such as an organ) whose local intensity distributions – given

by the local RIQFs – are distinguishable from those of neighboring volumes. The local RIQFs serve as samples from these large-scale volumes. The local RIQFs were clustered, with each resultant cluster population representing such a volume. In contrast to the local-clustered model, which only measured the variability in the local RIQFs within such a volume, the combined-clustered model combined the local RIQFs (and thus the local regions) belonging to a volume within each image and measured the RIQF variability across training of these larger-scale volumes.

One deficiency of both the local-geometric and combined-clustered schemes is that the RIQF models in the exterior of the object are static relative to the object geometry: at any particular position the RIQF model asserted is always the same. Yet the image regions that these models are supposed to characterize are found at different object-relative positions in different images. Figures 5.1 and 5.2 show the motion of the regions on the exterior of the bladder for two patients. For example, the region representing the neighboring pubic bone may change position relative to the bladder surface across days, as the bladder fills and empties or as gas and fecal matter in the rectum provoke some change in the relative position of the bladder. These changes clearly affect the bladder-external appearance and thus how segmentation of the bladder might proceed. However, the correspondence provided by the m-rep model is determined relative only to the organ of interest and not those neighboring organs and volumes that affect the external appearance. The result is a false association of bone type and non-bone type distributions based upon geometric correspondence that is a source of confusion in the appearance model. While bone intensities above some threshold are treated separately (Section 4.5.1), the bone-type intensity distribution itself consists of a broad range of intensities, including some not observed elsewhere relative to the bladder (see Figure 5.1).

The deficiency of object-based correspondence arises due to two factors, the actual changes in the relative positions of objects and the correspondence between the atoms

of instances of an m-rep shape model. Regarding the first factor, in addition to the bladder-bone example above, neighboring organs such as the bladder and prostate may also be connected structurally, but connected at different places relative to each other in different patients. This may result in an image correspondence that is incompatible with the geometric correspondence if the single-object geometry does not model or account for this connection. Regarding the second factor, the atom correspondence between deformations of an m-rep is carefully manipulated to benefit *shape analysis* during the training of the m-rep shape model (Merck et al., 2008). Disregarding external appearance in this training is appropriate, but it contributes to a lack of image correspondence between geometrically corresponding regions (Section 2.4).

One solution to this general problem of correspondence that is deficient for external appearance is shape-based, having the shape model account for the changing conformation of the objects. Diffeomorphic warps are an example of such a shape model, whereby some template or probabilistic atlas of the 3D volume encompassing the organ of interest and its neighborhood is warped to fit the target image and the segmentation of the template is carried over (Davis, 2008). However, the motion of one organ sliding along another is non-diffeomorphic. In that such motion is a source of changing object-relative external appearance, the warping techniques are not applicable.

Other examples of shape models that account for changing external conformation are multi-object shape models, such as the grouped level-set methods of (Rousson et al., 2005; Tsai et al., 2004) and the multi-object m-rep segmentation framework of (Pizer et al., 2005b). These methods often capture the variability of the object conformations. During multi-object segmentation, the shape model consists of a current estimate for each object. In that the neighboring volumes of interest are other objects within the multi-object complex – like bone, rectum, and bladder to the prostate – or interstitial tissue defined to be outside of all the objects, the intensities within each volume are easily available. A concern in regard to these techniques is the complexity of the multi-object

model versus the single object model. The accurate segmentation of one object may not require and may be adversely affected by the concurrent segmentation of all neighboring objects. I avoid the complexity of modeling the 3D shapes of the volumes and for my appearance models consider them only as they affect the local and larger-scale intensity distributions near the boundary of the object of interest.

In this chapter I address this deficiency of static object-relative RIQF models through an image-based rather than a shape-based approach. I present two novel image matches both of which allow *shifting* of the image models on the boundary to account for the variable object-relative positions of image regions. Both image matches require an optimization to compute the match of a prospective segmentation, within the overall posterior optimization in the shape space (see Section 1.2).

Allegiance-based shifting

The shifting process in both model presented in this chapter is implemented using the idea of allegiance. In both models, the exterior appearance is captured by many small or local regions around the boundary of the object, while relatively few larger-scale RIQF models describe the variability of combinations of these local regions (see Section 4.6.2 for how to combine the RIQFs of local regions). Each local region can be thought of having an allegiance, which is its assignment to one of the larger-scale regions: the larger-scale combinations are formed according to this allegiance. The set of allegiances of all points colors the surface and forms the regional delineation on the boundary, as seen in Figure 5.1. Shifting is accomplished through the changing allegiance of the local regions.

The shifting works to improve the fits of the larger-scale RIQFs with respect to the associated RIQF models and proceeds by iteratively changing the allegiance of local regions when the change leads to an improvement in the large-scale RIQFs. Figure 5.3 shows these iterative improvements to one of the large-scale RIQFs exterior to the

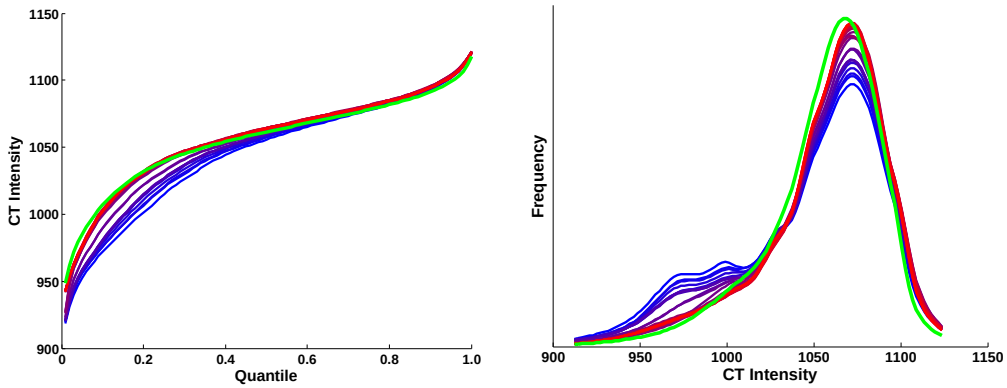

Figure 5.3: Series of large-scale RIQFs (blue to red) captured during shifting showing the improvement for the prostate-like volume exterior to the bladder. The mean RIQF of the volume over all training cases (the goal of the shifting) is shown in green.

bladder in the combined-clustered scheme, in this case representing the prostate volume. All the local regions on the surface are considered during each iteration, with a random order. The shifting requires the use of the neighbor relationships between the local regions, which is provided by the m-rep model (Section 2.4). If the allegiance of a local region is the same as that of all four neighbors, this local region is not on the boundary of the regional delineations and is thus skipped. For a local region that is on the boundary, the method considers the list of possible allegiances for this local region, which is the union of the allegiances of the neighboring local region and the current allegiance of this local region. An allegiance to one of the larger-scale regions is then chosen for the local region among this list of possibilities. This choice is made according to how the inclusion or exclusion of the RIQF of the local region improves the fits of the larger-scale RIQFs (see Section 5.3).

Two models: shifting-combined-clustered and shifting-local-geometric

The first model presented below, shifting-combined-clustered, extends the combined-clustered model (Section 4.6). Training of this model begins with the result of clustering on the local RIQFs. The cluster-type delineation on the object boundary, in which

every local region is assigned a cluster, implies larger-scale regions constructed from the cluster populations. This regional delineation is then deformed using the above shifting of allegiance, which allows particular local regions to change allegiance according to the image data.

The goal of the shifting is to improve the agreement with training of the larger-scale regions. Given only the rigid geometric correspondence, the larger-scale RIQFs formed from the initial allegiance can be a poor fit to the RIQF model of the true neighboring regions seen in training, due to the false associations implied by that correspondence. Allowing a shifting of the regional delineation thus leads to a more specific appearance for each larger-scale region. With the inclusion of a prior on the cluster-type per local region, this same deformation process is applied at target time to provide the optimization with an image match that reflects knowledge of the regional conformations seen in training.

In the second model, shifting-local-geometric, I extend this idea to the local-geometric image match (Section 4.4.3), which has so far led to the best results (Section 4.6.4). The extension can be thought of as the first model, but with the clusters chosen to be the RIQFs from each local region rather than the result of clustering on the RIQFs over all local regions. Again the goal in training is to allow the regions to deform to tighten the regional RIQF distributions. The local regions are subdivided into many subregions that can change allegiance to affect the sampling of a local region relative to the geometry. Given the small size of these very local regions and the fact that the RIQFs of neighboring regions at such small scale are often very similar, the deformation is more highly constrained.

The remainder of this chapter is organized as follows. Section 5.3 details the shifting-combined-clustered model. Section 5.4 details the shifting-local-geometric model. In Section 5.5 I present results testing the efficacy in segmentation of these models in comparison to their static counterparts, for bladders and prostates in CT. In Section 5.6 I conclude and motivate the future research and application directions described in

detail in Chapter 6.

5.3 Shifting Combined-Clustered Regions

In this section I describe the shifting-combined-clustered image match model in training and segmentation. The goal is a model that accurately represents the RIQF variability in the larger-scale volumes surrounding the object of interest, while also allowing these volumes to shift with respect to the object in order to reflect the different regional conformations seen across training images. By using shifting to move from the initial conformation closer to each training case, the algorithm that I present here accomplishes both goals.

Training of the shifting-combined-clustered model begins with the result of the combined-clustered model (Section 4.6.3), namely a set of several large-scale RIQF models representing the neighboring exterior volumes and a fixed allegiance of the local regions on the surface making up these larger volumes. For each training image, the allegiances of the local regions are changed according to the shifting scheme described in Section 5.2 and in more detail below. The result is an allegiance for each training case that is closer to the ideal but achievable from the initial allegiance. PCA is performed on the large-scale regions implied by these new allegiances to obtain a new set of large-scale RIQF models. Additionally, the final allegiances of each local region over all training cases are stored for use as prior at target time, completing the training of this model.

I will now present the training in more detail. Over all training images I_i, I compute RIQFs $\mathbf{q}_{ir}$ for many local exterior regions anchored to individual geometrically defined points on the object boundary indexed by r. Gath-Geva clustering on the pooled set of RIQFs over all positions and images results in cluster populations that can be thought of as samples from the neighboring volumes observed in the training images, one volume per cluster population (Section 4.6.1).

Each RIQF $\mathbf{q}_{ir}$ and its associated local region r in image i is then assigned a cluster that maximizes the class probability $h(k|\mathbf{q}_{ir})$ determined by the Gath-Geva technique. This results in a partition of the object boundary which is different for each image depending on the different relative positions of the neighboring volumes, as seen in Figure 5.1. The method then combines the local RIQFs of each cluster type, and thus volume, in each image to obtain the K sets (one per cluster) $\{\mathbf{q}_i^k, \forall i\}$. Principal component analysis on each of the K sets results in a Gaussian model of the RIQF variability for each of the neighboring volumes observed across training.

The next step in the training process is the shifting of the initial allegiance (cluster assignment) in each of the training cases. The goal of the shifting in training is to obtain the best regional delineation achievable from the initial allegiance. This is due to the reasoning that one should do the same thing in training as at target time and that at target time the model will not be at the correct answer as it is in training. Therefore, the shifting must start at a reasonable regional conformation. The modal allegiance is such a conformation. The modal allegiance of a local region r is given by the mode over the training images of the allegiance of r, which is determined by the class probabilities $h(k|\mathbf{q}_{ir})$ from above.

The general scheme for the shifting is described in Section 5.2 above. There are two aspects to determining a new allegiance for the local region r related to which large-scale regions benefit from the inclusion of the RIQF at r. Consider the local RIQF $\mathbf{q}_{ir}$ at region r, and let the current large-scale RIQFs in this image be given by $\{\mathbf{q}^k, \forall k\}$. If p is on the boundary of regions j and k and is aligned with cluster (and thus large-scale region) k, there are two questions that are balanced in determining whether to change the allegiance of p. First, how much does including $\mathbf{q}_{ir}$ in $\mathbf{q}^j$ improve or make worse the match of $\mathbf{q}^j$? The second question is how much does excluding $\mathbf{q}_{ir}$ from $\mathbf{q}^k$ improve or make worse the match of $\mathbf{q}^k$? Regarding these two questions, if the net result of changing allegiance is an improvement, the change is made.

The inclusion and exclusion are done using the mixing function $\oplus$ detailed in Section 4.6.2. Let w_{ir} be the number of voxels contributing to $\mathbf{q}_{ir}$, and w^k and w^j the number of voxels contributing to the large-scale regions k and j respectively. See Section 4.4.1 and Equation 4.4 for computing the penalty of an RIQF with respect to some PCA model on RIQFs. The first aspect of the allegiance at r, whether including $\mathbf{q}_{ir}$ in $\mathbf{q}^j$ improves $\mathbf{q}^j$, is more formally stated as an inequality:

$$-\log p\left(\mathbf{q}^j \oplus \frac{w_{ir}}{(w_{ir}+w^j)}\mathbf{q}_{ir}\right) - \left(-\log p(\mathbf{q}^j)\right) < 0. \tag{5.1}$$

If the inequality is true, including $\mathbf{q}_{ir}$ in $\mathbf{q}^j$ leads to $\mathbf{q}^j$ being closer to the mean of the RIQF model for the j^{th} large-scale region. I have left in the double negation to keep clear the relationship with Equation 4.4. The second aspect, whether excluding $\mathbf{q}_{ir}$ from $\mathbf{q}^k$ improves $\mathbf{q}^k$, leads to another inequality:

$$-\log p\left(\mathbf{q}^k \oplus \frac{-w_{ir}}{(w^k-w_{ir})}\mathbf{q}_{ir}\right) - \left(-\log p(\mathbf{q}^k)\right) < 0. \tag{5.2}$$

Let the left-hand sides of Equations 5.1 and 5.2 above be referred to as $\Delta logp^j$ and $\Delta logp^k$ respectively. They measure the proposed penalty minus the current penalty for large-scale regions j and k, in terms of squared Mahalanobis distance. If the current penalty is larger, the inequality is true and the associated large-scale RIQF would be improved by the proposed change.

In deciding whether to change the allegiance of local region r, the shifting scheme takes both $\Delta logp^j$ and $\Delta logp^k$ into account. In that the same number of modes are kept for each of the regional RIQF models, the units of the two penalty differences are the same, so their simple sum gives an indication of whether the change in assignment is beneficial. Additionally, the shifting accounts for the number of voxels contributing to each region, making the same improvement in a larger region more important than in a smaller one. Overall, a score is computed for each large-scale region j that borders the

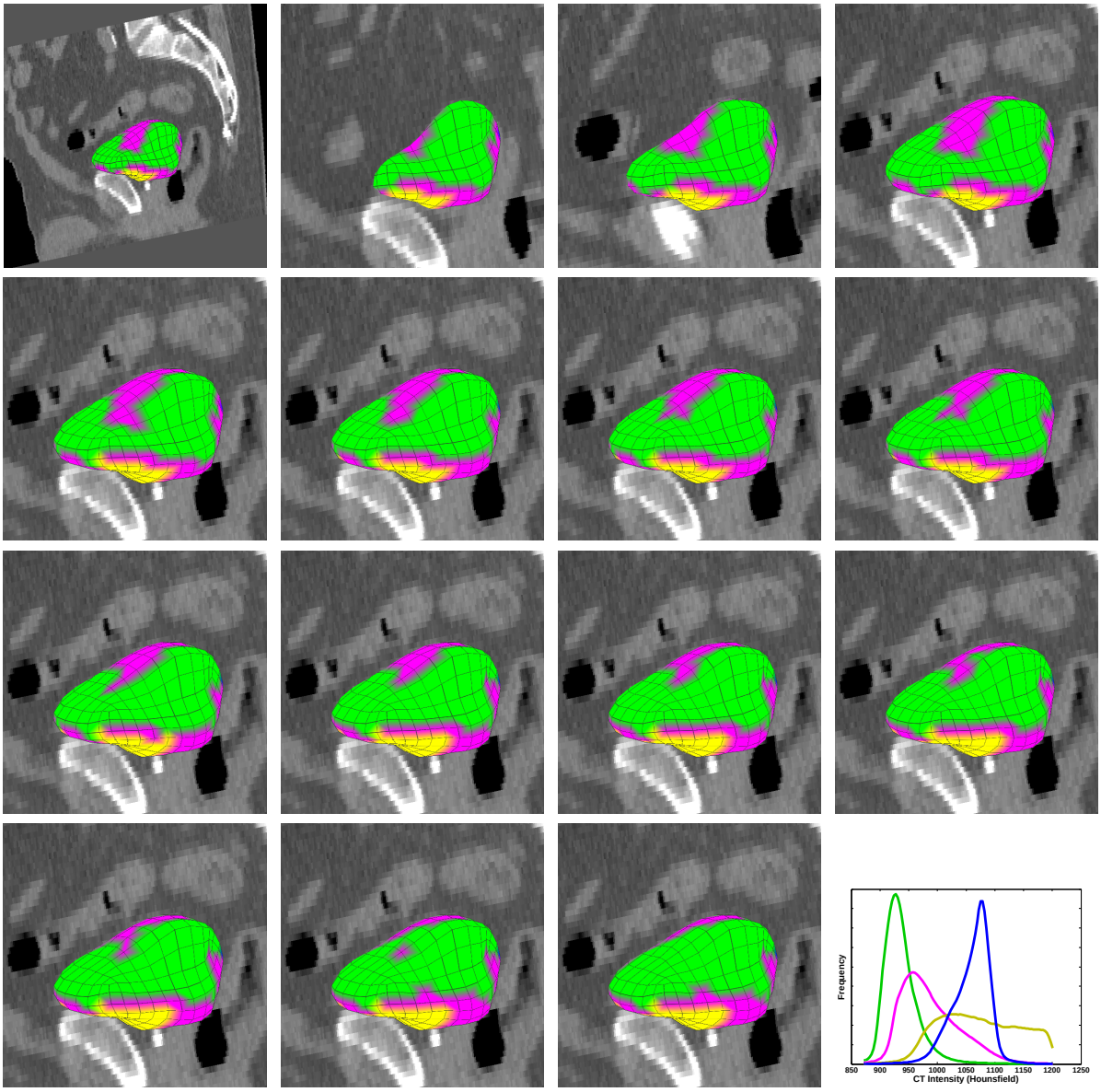

Figure 5.4: Series of images showing the shifting of the modal regional delineation required for a particular image. The colors represent the different cluster assignments, made per point, and the associated large-scale intensity distributions are shown in the bottom-right frame. Top row: image slices relevant to the part of the surface shown. While the initial delineation expects some of the magenta region on top, the particular image does not support this.

point r:

$$changeScore_r(j) = (w^j + w_{ir})\Delta logp^j + (w^k - w_{ir})\Delta logp^k \tag{5.3}$$

For $j = k$ this *changeScore* is set to zero, since the large-scale RIQF $\mathbf{q}^k$ already includes

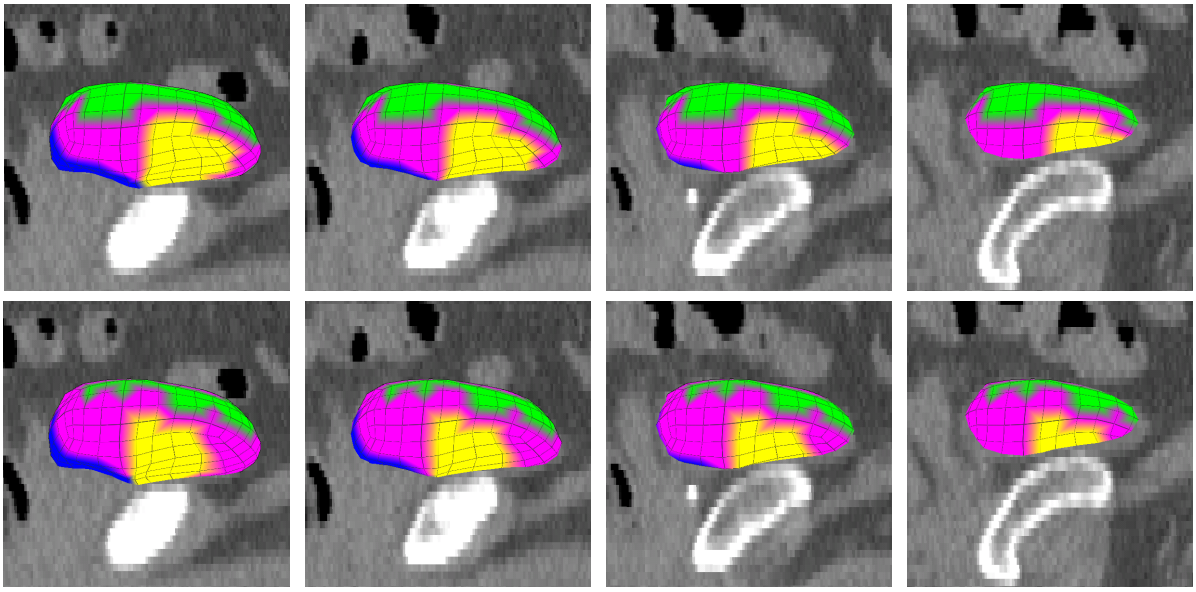

Figure 5.5: Showing the result of shifting on the exterior of the bladder. Top row: initial regional delineation. Bottom row: the delineation after shifting. Note that the position of the region corresponding to the pubic bone has changed to better fit the image.

the contribution of $\mathbf{q}_{ir}$ and thus cannot include it again. The scores for all bordering regions are compared and the allegiance of r is changed to the large-scale region with the minimum score.

The shifting procedure repeatedly iterates over all points on the boundary until there is no change of assignment or a few points oscillate in their assignment. The oscillation happens, for example, when two large-scale regions j and k are improved by the same local region r on their shared boundary. If r is currently aligned with k, then $\Delta logp^j$ is multiplied by $(w^j + w_{ir})$, which gives more weight to the improvement associated with changing the allegiance of r. Once r is aligned with j, this additional weight is again given to changing the allegiance, this time back to k.

An additional constraint in the shifting is that a singleton local region is not allowed: if r does not share its allegiance with any of its neighbors, its allegiance is forced to change. This keeps the regional delineation in step with the idea of large-scale neighboring volumes. Figure 5.4 shows the shifting procedure on the superior surface of the

bladder. As seen in the top row, at first, the modal configuration includes some brighter (magenta) regions superior to the bladder. However, this configuration does not fit the particular image data, which contain fat and bowel in that area. The series then shows the changing cluster assignments that lead to the regional conformation evident in this image. Figure 5.5 shows before and after shifting for an additional image.

Each training image undergoes the same shifting procedure, which results in a new set of large-scale RIQFs observed over training, $\{\mathbf{q}_i^k, \forall i\}$, for each cluster k. These samples reflect the large-scale RIQFs given the regional configurations that are achievable through shifting from the modal configuration. PCA on each of the K sets results in a Gaussian model of the RIQF variability for each of the neighboring volumes observed across training. Additionally, a prior is computed reflecting the relative frequency of each cluster assignment per point, stored as a $K \times N$ matrix, where N is the number of points and the sum of each column is one. The presence of gas and bone are accounted for per large-scale region, as in the combined-clustered model (Section 4.6.4). At this stage, training is complete.

At target time the goal is to return an image match that reflects the large-scale RIQF match, given a likely regional conformation relative to the object. The image is sampled relative to a prospective segmentation to populate the set of local RIQFs. The local RIQFs are combined into the large-scale regional RIQFs according to the mode of cluster assignments per patch, as in the combined-clustered model. The large-scale RIQFs and the mode of the cluster assignments, which leads to the modal regional delineation, are the inputs to the shifting procedure. The shifting procedure is as described in training with the addition of the prior on cluster assignment at the level of computing $changeScore_r(j)$ in Equation 5.3. Specifically, $changeScore_r(j)$ is modulated by the prior probability of the point r being assigned the cluster j, which is simply the relative frequency of that allegiance observed over the training cases that was recorded above. The prior also dictates which clusters can even be considered at the point r, which

restricts the shifting to conformations that were seen in training at the local scale. The image match returned for the prospective segmentation is then, after shifting, the sum of the squared Mahalanobis distances of the large-scale RIQFs in the PCA spaces defined by their respective RIQF models.

5.4 Shifting Local-Geometric Regions

This section describes the shifting-local-geometric image match model, which extends the local-geometric match of Section 4.4.3. The impetus in general behind shifting of the exterior image model on the boundary is to relax the rigid correspondence defined by the geometric model in order to achieve tighter regional image statistics and as a result a better image-implied correspondence. This idea was originally applied to the large-scale exterior volumes and their changing conformations; the shifting-combined-clustered model from the previous section implements shifting on those regions. In this section I extend this idea to the smaller-scale regions of the local-geometric model. I will describe the training of this shifting-local-geometric model and the target time computation of image match. The scheme is analogous to that in the previous section, but at a smaller scale and with stronger constraints on the shifting.

The local-geometric image match of Section 4.4.3 consists of an RIQF model for each of many local regions anchored to points on the surface of the m-rep. This is a very specific appearance model, as the RIQF model computed at any point relies only on the image data seen at that corresponding point across training. The m-rep model supplies explicit correspondence between the training cases (Section 2.4).

However, in relying solely on the geometry-supplied correspondence, some local regions are not properly modeled. For example, consider a part of the boundary abutting some organ in one image and fat in another; over many training images, the local region encompassing this part of the boundary will contain varying mixtures of the associated

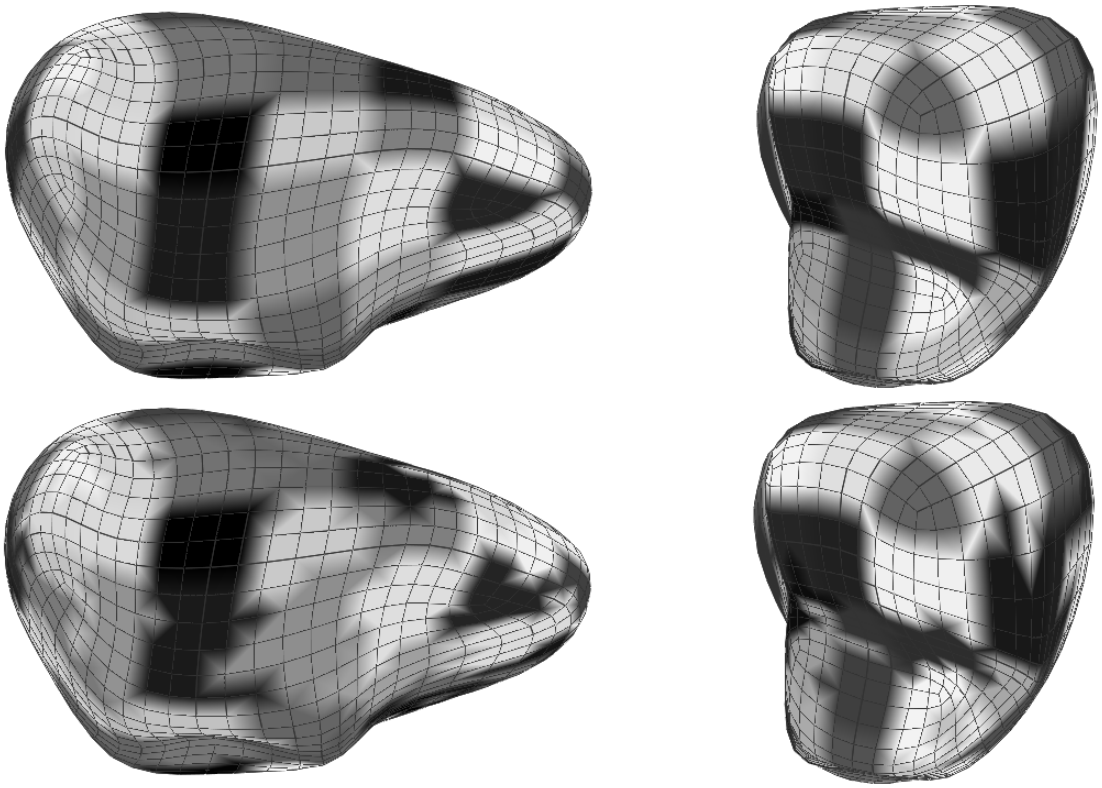

Figure 5.6: Top: the starting region assignment per subregion for the shifting-local-geometric model. The grid pattern shows the anchor points of the subregions while the color shows the associated local region. Bottom: after shifting.

organ-type and fat-type intensity distributions. These changes in mixture are not modeled linearly in the RIQF feature space, so PCA will not appropriately capture this variability (see Section 2.3.1). Thus the RIQF feature space and linear statistics on it should not be responsible for capturing these changes. I propose some small shifting of the region along the surface to reduce the effect of these changes in the RIQF space. The shifting procedure is the same as that presented in Section 5.3, but on smaller regions, with the goal to improve the fit of each local region RIQF to its respective RIQF model.

To implement this small-scale shifting, I subdivide each local region into many smaller subregions – on the order of 16 per local region[1]. Each subregion is originally assigned to the local region it is encompassed by, leading to the delineations shown in the top row of Figure 5.6. The local region RIQFs are constructed by combining the RIQFs of the associated subregions (Section 4.6.2). PCA on the set of local region RIQFs at a given point r across training leads to an RIQF model for r. Each of the local regions are treated separately in this way, leading to a starting RIQF model per local region. These models are equivalent to the models used in the local-geometric match (Section 4.4.3).

The shifting procedure iterates over the subregions on the surface, changing the allegiance of the subregions in order to improve the fit of the observed local RIQFs to the RIQF models. In a given training image the initial delineation on the fitted model surface is as shown in the top row of Figure 5.6. In the following, let s index the subregions, and let k index the local regions on the surface. k_s denotes the local region currently associated with the subregion s, $k_s \in \{1, \ldots, K\}$. In the experimental results presented below for example, K is 78 for the bladder and 74 for the prostate, the number of local-scale regions making up the surface, while the number of subregions is 1218 and 1154, respectively. Further, let $\mathbf{q}_s$ denote the RIQF of subregion s and $\mathbf{q}^k$ denote the RIQF of the local-scale region k, which is computed by combining the RIQFs of the set of subregions aligned with local region k, $\{\mathbf{q}_s, k_s = k\}$.

Analogous to the shifting-combined-clustered method, the local region allegiance of subregion s may change according to how much including $\mathbf{q}_s$ in $\mathbf{q}^j$ improves the match of $\mathbf{q}^j$ and how much excluding $\mathbf{q}_s$ from $\mathbf{q}^k$ improves the match of $\mathbf{q}^k$ (Section 5.3). The

[1]The surface of the m-rep model is implied by the finite set of spoke ends of the medial atoms (see Section 2.4). Subdivision is used to determine the surface at interpolated positions, and each successive subdivision process roughly quadruples the number of points on the implied surface. Thus, 2 levels of subdivision on the surface of the m-rep model provides the subregions each with approximately $1/16^{th}$ the area of the full regions. The number of subregions was chosen heuristically to ensure that subregions are larger than voxel-scale

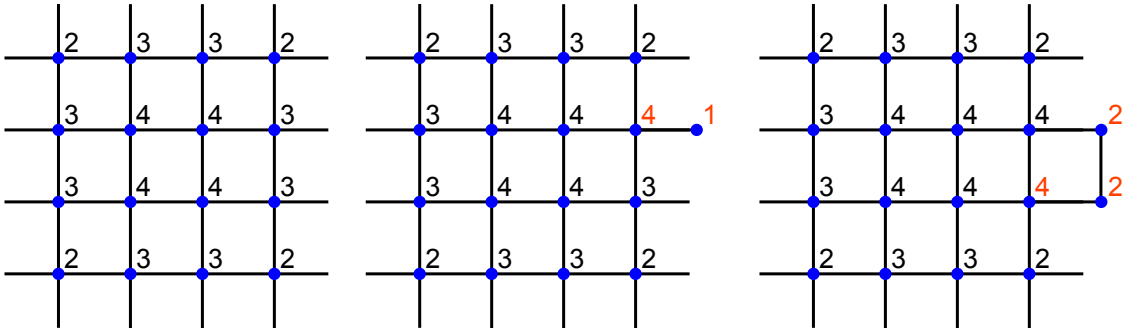

Figure 5.7: Showing the adjacency of a local region changing according to the subregion allegiance. Left: the original region with total adjacency of 48 with 16 subregions, for an average adjacency of 3 for the whole region. Middle: a neighboring subregion changes allegiance, adding 2 to the total adjacency. Right: an additional neighboring subregion changes allegiance, adding 4 to the total adjacency.

energy differences $\Delta logp^j$ and $\Delta logp^k$ are computed as in Equations 5.1 and 5.2, where rather than a local region contribute to a large-scale region, a subregion contributes to a local region.

In addition to the sum of $\Delta logp^j$ and $\Delta logp^k$ in determining whether to change the region assignment of s, I add a connectivity constraint on the subregions making up a local region. This is to ensure that a region remains "region-like," as in the initial delineation. The connectivity constraint is necessary due to the fact that neighboring subregion RIQFs may be very similar to one another, in that they are constructed using a limited number voxels and that they may share some contributing voxels. Additionally, neighboring RIQF models may be similar, particularly when their associated local regions abut the same large-scale exterior volumes.

The connectivity constraint consists of a penalty against changing the average in-region adjacency. The in-region adjacency of a particular subregion is the number of neighboring subregions with the same allegiance. The subregion neighbors are given by the quad-mesh of the m-rep surface (see Section 2.4). The average adjacency of a local region is the sum of the in-region adjacency over all the constituent subregions – which is the total adjacency – divided by the number of subregions. For example, consider a

4×4 grid of subregions making up a local region, as in Figure 5.7. The interior four subregions have neighboring subregions all assigned to the same region, for an adjacency of 4. The 8 side subregions have three neighbors of the same type (adjacency 3), and the 4 corners have in-region adjacency of 2, giving the collection an average adjacency of 3. Adding a subregion to the local region (changing its allegiance to the local region) will add 2 to the total in-region connections and 1 to the total number of subregions, changing the average adjacency to 2.94. Adding a second subregion, connected to both the original grid and the new addition, will add four total connections and one subregion, changing the average adjacency back to 3.

To constrain the shifting motion, I construct a Gaussian on this average adjacency, using the heuristically obtained standard deviation of 1/15 of the adjacency of the initial region. I tried many standard deviations for the average adjacency term, from 1/2 to 1/30, with the result being more or less deformation allowed for the local regions. No deformation, which generally happened at 1/30, simply corresponds to the static local-geometric image match and resulted in segmentation accuracy equal to the result presented for that image match. I considered too much deformation to be when local regions became disconnected in training. The value 1/15 was chosen as the standard deviation because that value allowed the most deformation without disconnected regions in training.

The connectivity constraint is, as with the other terms, a difference in the squared Mahalanobis distance on the adjacency of both regions j and k before and after the proposed change in assignment. As $\Delta logp^j + \Delta logp^k$ determines how much the proposed change in assignment improves the image match of the two regions, the connectivity constraint determines how much the proposed change improves the shape of the regions relative to their initial shapes.

This shifting procedure is applied to each training case, resulting in shifted regions and new local RIQFs $\mathbf{q}^k$ for each image. The bottom row of Figure 5.6 shows the

subregion assignments after shifting on an example case. PCA on the corresponding shifted local RIQFs over training leads to an RIQF model per local region, as in the local-geometric model. However, in that the shifting is defined to reduce the RIQF variability, these RIQF models have tighter variances than those of the local-geometric image match. Additionally, a prior is computed for each subregion on the region assignments found in training. This completes the training of the shifting-local-geometric model.

At target time, the image match of a prospective segmentation proceeds as follows. The subregion RIQFs are sampled relative to the geometry and combined to form the local RIQFs using the initial delineation described above. The shifting procedure is then applied as in training, with the addition of the prior per subregion computed in training. This prior serves only to discard region assignments never observed in training. The shifting procedure moves the regional delineation to best fit the data. After shifting, the image match returned is the sum of the squared Mahalanobis distances of the local RIQFs in the PCA spaces defined by their respective RIQF models.

5.5 Segmentation Results

In this section I describe an experiment testing the efficacy in segmentation of the shifting-combined-clustered and shifting-local-geometric image matches. Automatic segmentations of the bladder and prostate in within-patient image sets were generated via posterior optimization of m-reps, as detailed in Sections 1.2 and 2.4. The image matches are interchangeable in the optimization, allowing me to compare the efficacy of the various matches with all else as constant. I compare the results obtained using the matches described in this chapter to those obtained using the matches' static counterparts of Chapter 4, the combined-clustered and local-geometric models.

The experimental setup is as described in Section 4.5. The data consist of 5 patient pelvic CT image sets, each set containing approximately 16 daily images, and each

image having corresponding manual segmentations of the bladder and prostate for use in training and in comparison with the automatically obtained results. I performed leave-one-day-out training for a given patient, with testing on the left-out day, for each of the 5 patients independently. I compared the automatically obtained results to the manual segmentations via the Dice similarity coefficient (DSC), average surface distance (ASD), and Hausdorff maximum surface distance (MAX) measures, which are computed using the compare_byu tool described in Section 2.5.

The initialization of all the tested methods was the result of the automatic segmentation using the local-geometric model. This is justified in that first, the goal is to improve upon the best results, and second, the shifting matches of this chapter are inappropriate far from the right answer. At target time, a prospective segmentation is probably poorly aligned with the true organ boundary. While with a static model this will lead to a poor match, after shifting the match may be greatly improved. The shifting image matches return the best image match among possible regional conformations (those seen in training). Consider for example a bladder image where the prospective segmentation is smaller than the true bladder on the superior side. The intestines, with similar intensity to bladder, occasionally abut the superior side of the bladder (as seen in the top row of Figure 5.4). The shifting match would reflect that possibility and thus mute the large penalty of a static match, returning a match reflecting the presence of intestines superior to the model. Thus the shifting match is not appropriate far from the right answer.

To account for this, I ran the segmentations using all four image matches, but rather than start at the mean shape initialized using the bone transform, I ran the optimizations from the result of segmentation using the local-geometric match, which has so far led to the best results. With the prospective segmentations closer to the right answer, the shifting has less opportunity to confuse the optimization.

Figure 5.8 shows the median volume overlap, average surface distance, and maximum

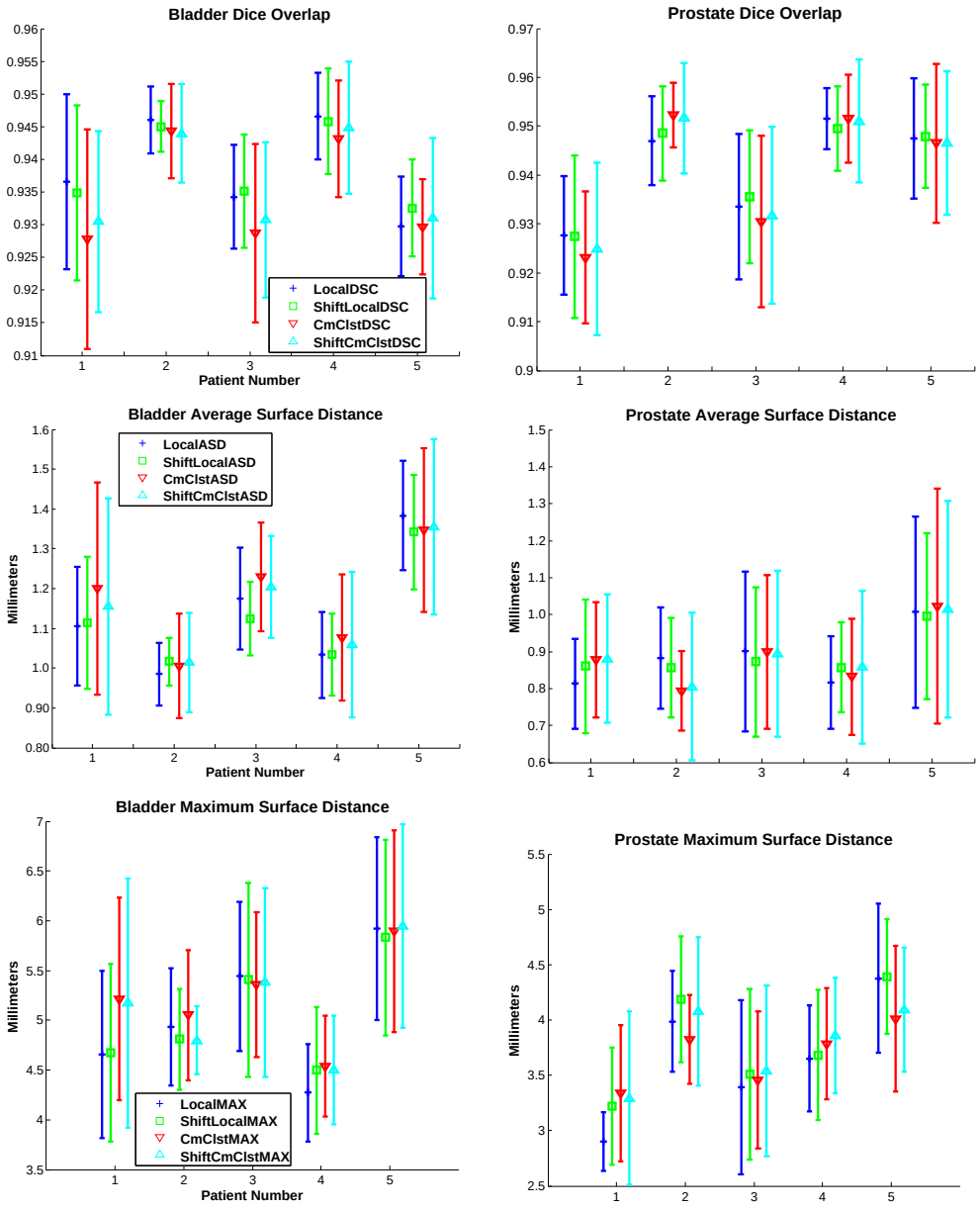

Figure 5.8: Experimental results comparing the four appearance models. For each patient and model, the median is plotted with the 95% confidence interval. Top row: Volume overlap results on bladder (left) and prostate (right). Middle row: average surface distance. Bottom row: maximum surface distance.

surface distance measures per patient for each of the four appearance models, local-geometric (Local), shifting-local-geometric (ShiftLocal), combined-clustered (CmClst), and shifting-combined-clustered (ShiftCmClst). For each patient and model, the median and 95% confidence interval are plotted to show the spread of the results.

The shifting-combined-clustered model led to slightly improved results over the combined-clustered model. Regarding the bladder, the shifting approach proved better than the static in a majority of the images for 4 of 5 patients, with respect to both overlap and surface distance. For prostates, this was true in 3 of 5 patients. These results are not statistically significant. They do however imply that shifting of the large-scale regions can help segmentation accuracy and usually does not hurt it.

However, even with shifting, a match based on a few large scale regions does not have the efficacy of a match based on more local regions. The static and shifting local-geometric models led to improved bladders over the large-scale matches, with respect to overlap and surface distance, in 71%-76% of the images (55 to 59 out of 78 total). The more local matches were better in prostate volume overlap in 70% of images.

Finally, the shifting-local-geometric image match proved better than its static counterpart in bladder segmentations but not in prostates. For bladders, the shifting scheme led to improved volume overlap in 58% of cases overall, and improved a majority of a patient's images in 3 of 5 patients. Additionally, in images where local-geometric model proved better, it was better by less than when the shifting approach was better (averages of 0.20% and 0.23% volume overlap respectively). For prostates, the local-geometric match led to the best overlaps in 60% of cases overall and 4 of 5 patient image sets.

Not shown in the tables, I also ran a global RIQF match segmentation from the local-geometric result. This yields negligible change, which implies that the local-geometric result, which is at a relative minimum of its objective function, is also at a relative minimum of the objective function using the global RIQF match.

I have rerun these experiments a number of times with various changes in the training. For the shifting-combined-clustered model, one such variation is in the computation of the *changeScore*. In Equation 5.3 the sum of $\Delta logp^j$ and $\Delta logp^k$ is weighted by the size of the competing large-scale regions. I have also used the unweighted sum and the weighted average instead. These variations led to no change in the overall trend.

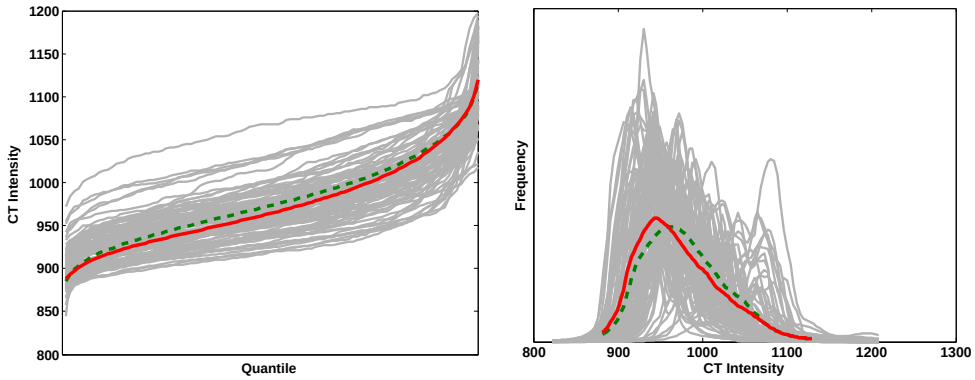

Figure 5.9: Left: In grey, the many constituent local RIQFs that make up a large-scale region exterior to the bladder. In red (solid), the combined RIQF, which is close to the mean of the region's RIQF model, shown in green (dashed). Right: Associated densities.

Another variation addressed the issue in shifting of combining RIQFs that are too different. One can combine disparate RIQFs that are each a bad fit with respect to an RIQF model, and yet the combined RIQF may be a better fit than any of constituent RIQFs. Figure 5.9 shows this in a simple example, where the RIQFs of many local regions are combined to form the large-scale RIQF representing the magenta region of Figure 5.1. The combined RIQF fits the model better than the constituent RIQFs. During the shifting, a very different RIQF may be combined with a larger region, if the combination has a better fit than without the different RIQF. However, if I consider a large-scale region to have similar local distributions as noted in Section 5.2, then RIQFs that are *too different* must belong to different large-scale regions and so should not be combined.

I addressed this issue by adding an additional penalty against disparate RIQFs being included in the same large-scale region. To define the penalty, I applied PCA to the set of RIQFs belonging to a cluster over the training cases. The resulting RIQF models are those from the original local-clustered scheme of Section 4.4.2: the RIQF models measure the variability in the local RIQFs within a large-scale region. The additional term in the *changeScore* penalized adding a local RIQF to a large-scale region when that

RIQF was not like the constituent local RIQFs already in the region. While ensuring the homogeneity of the large-scale regions may be appropriate, this additional term did not affect overall results.

While the results show some differences between the various methods, they are all within the context of highly accurate segmentations. The fits computed with respect to the expert segmenter, regardless of the particular image match method, often exceeds the agreement found *between* experts in our own studies.

Comparing my results against the best obtainable result rather than against the expert result

A confounding factor in these results that may limit both the overall efficacy observed and the differences observed between the methods is as follows: the automatic segmentations for any particular image are all limited by the optimization to be within the "shape space," and further, this space *may not* include the exact expert segmentation of the image that the automatic results are compared against. In this section I present the results of an additional experiment in which I first computed the segmentation that is both closest to expert's manual result *and in* the shape space, and then compared my automatically-obtained shifting and clustered results against this "best obtainable" segmentation.

The best obtainable segmentation is computed using the same shape space as my results while also using the binary expert segmentation rather than the original CT as the image data. This allows for the use of an image match based on the shape model's surface distance to the unambiguous boundary of the binary image. The fitting of the m-rep model to binary data is a separately developed tool that is described in (Merck et al., 2008) and (Han et al., 2007). Rather than initialize the deformable model fitting at the mean model in the shape space (see Section 1.2), I use the projection in the shape space of a previously fit model that had many more degrees of freedom in matching the

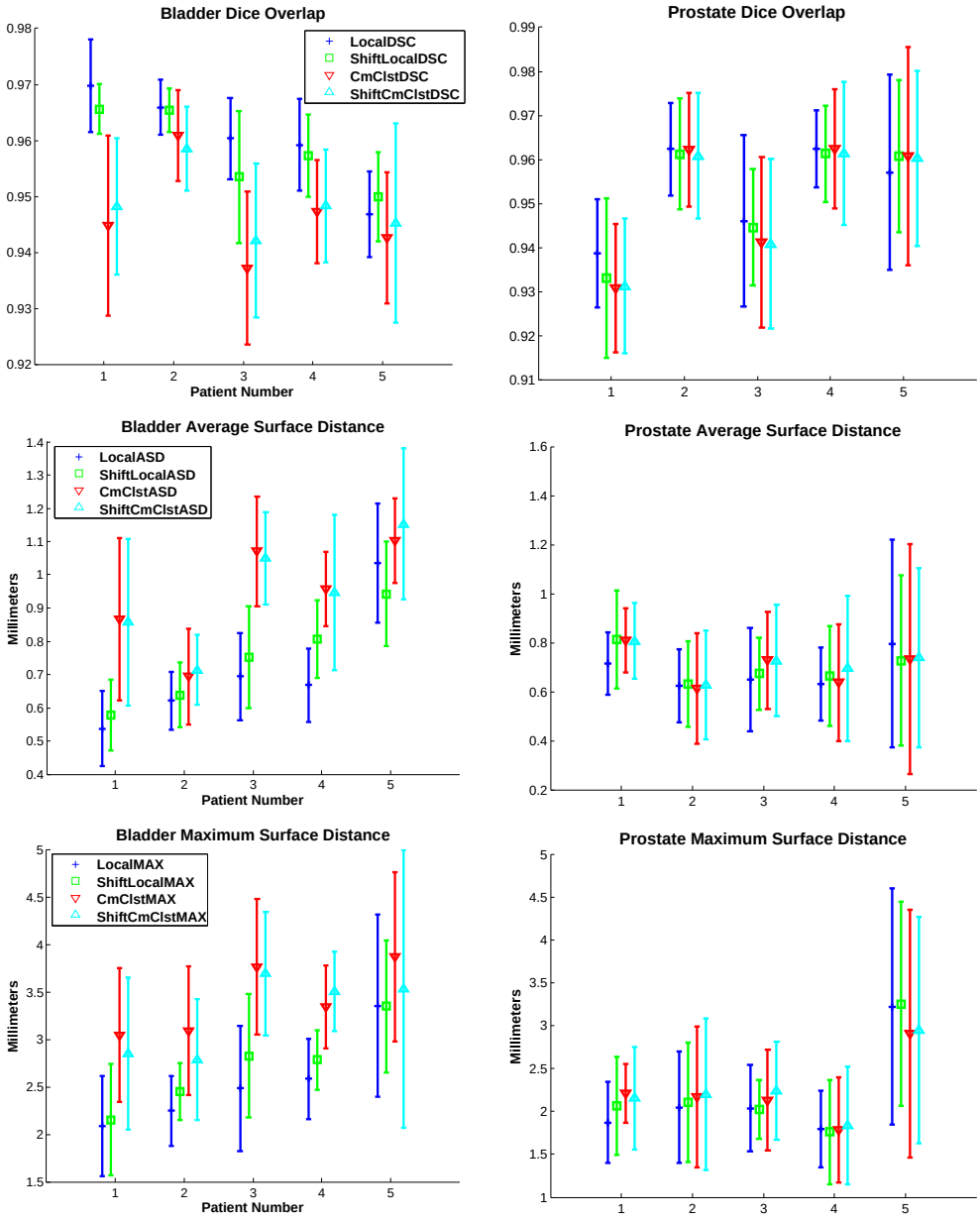

Figure 5.10: Experimental results in the vein of Figure 5.8 comparing the four appearance models, but against the best obtainable segmentations rather than the expert segmentations. For each patient and model, the median is plotted with the 95% confidence interval. Top row: Volume overlap results on bladder (left) and prostate (right). Middle row: average surface distance. Bottom row: maximum surface distance.

binary image surface (see (Merck et al., 2008)). I also downweight the shape prior term, which would force the model to stay close to the mean. The result of the deformable model fitting is a segmentation that is both as close as possible to the expert manual

segmentation and still within the shape space that constrains my shifting and clustered results.

Figure 5.10 shows the median volume overlap, average surface distance, and maximum surface distance measures per patient for each of the four appearance models when the comparison is made against the best obtainable segmentation. The results show the same trends as before, namely that shifting is often an improvement over static and that local regional image matches are improvements over larger-scale clustered regional image matches. Further, the results are widely improved over those shown in Figure 5.8, which compared against the exact expert segmentations. In particular the Hausdorff distance shows a marked improvement, from approximately 5 mm to 2.5 mm in the bladder cases for example. This leads to the conclusions that first, the shape space itself limits the possible fit against the expert segmentations and second, the shifting and/or clustered results approach this best possible fit.

5.6 Conclusions

In this chapter I have presented two appearance models that reflect the changes in large-scale or local-scale regional conformation seen in training, with the goal to improve automatic segmentation using the resulting image matches. If the outside appearance is modeled with RIQF models associated with object-relative positions, the two image matches allow the RIQF models to change their associated positions in ways that reflect the variability seen in training and that lead to improved fits with respect to those RIQF models. The first appearance model, shifting-combined-clustered, applied this idea to the large-scale regions of the combined-clustered match of Chapter 4. The second model, shifting-local-geometric, took as a starting point the model that led to the best results so far, the local-geometric, and applied this idea at that more local scale.

I tested these image matches with regards to their efficacy in deformable model

segmentation of bladders and prostates in CT. I found that shifting helps the larger-scale image match a lot. However, at the local-scale, shifting improved bladder segmentations but not prostates. And yet, these results are within the context of highly accurate segmentations, whose fit with respect to the expert segmenter exceeds the agreement found between experts in our own studies.

These results are encouraging and point to some future directions. First, the improvement in bladder segmentations suggests that when the appearance is more highly variable, shifting can improve the final fits when the model is initialized using a static match. This variability in the neighborhood relationships is more present across patients than within a patient (see Section 5.2). As the segmentations performed in this chapter were all within patient, it is expected that the benefit provided by shifting will be more clear in across-patient studies. Additionally, there are many other organs that share this variability in the external conformation, such as the hippocampus and caudate of the brain. The ventricle of the brain abuts these organs to varying extent across images; shifting may help account for this.

A second direction of research is in using the shifting of large-scale regions within the context of multi-object segmentation. I noted in Section 5.2 that I hope to avoid modeling the 3D complexity of the many objects and volumes affecting external appearance. However, in finding homogeneous regions relative to a particular object, and in modeling their changing positions, my appearance model does provide information about the location of neighboring objects. The degree to which the objects being modeled are consistent with the regional delineation on a particular object can be used as an image term in a deformation of the multi-object complex.

A third direction of research is in the basis of the shifting scheme. Currently the allegiance-based shifting is a discrete deformation, where a number of points at fixed positions can change allegiance to affect the overall conformation. In that the motion of the neighboring volumes being modeled may not align themselves perfectly with

the discrete shifting I have devised, there may be an advantage to a more continuous deformation of the region sampling. An example continuous deformation would be a diffeomorphic warp of the quad mesh on the surface of the object model based on spherical thin-plate splines (Zou et al., 2007). These and other directions of research are discussed in Chapter 6.

Chapter 6

Discussion and Future Work

In this dissertation I have presented novel image appearance models for use in automatic medical image segmentation via deformable models. These appearance models and their associated image matches are characterized by clustering and/or by shifting on object-relative image descriptors. I use clustering to decompose the image around the object of interest into larger-scale regions that are then characterized separately. I use shifting to account for variability in the position of regions. Further, I considered combinations of these techniques in the models I have developed here. Throughout, the efficacy of these appearance models has been shown within the context of Bayesian deformable model segmentation of CT images of the abdomen and male pelvis, with the organs of interest being the kidney, bladder, and prostate. In this chapter I conclude this work, first discussing my contributions and then suggesting future directions of research.

6.1 Review of the contributions of the research

This section reviews and discusses the contributions of this work as related to the thesis and claims specified in Chapter 1 (Section 1.4). I list each claim and summarize the evidence for it found in this dissertation. Finally, the thesis is reviewed.

1. *I demonstrate that clustering on local image descriptors can be used to construct effective appearance models for deformable model segmentation.*

The decomposition of an image into regions is accomplished through clustering. In this dissertation I have demonstrated the effectiveness of clustering-based appearance models by developing such models and testing their efficacy. I presented two appearance models that use clustering on object-relative image descriptors to characterize appearance changes near the object boundary. The first model was the intensity profile clustering of Chapter 3. This model used clustering on cross-boundary intensity profiles to create profile types that characterize boundary appearance. The associated correlation-based image match was tested on kidney segmentation in CT. The second model was the combined-clustered scheme described in Section 4.6. This model considered the variability in the appearance of large-scale regions neighboring the object of interest. This variability was captured via Gaussian models on RIQFs (Section 2.3.2). The combined-clustered model was tested on bladder and prostate segmentation in CT.

Both clustering-based appearance models showed advantages in segmentation accuracy over previous models that used either a constant or a global-regional appearance (Sections 3.4.1, 4.6.4). These advantages are discussed below in Claims 2 and 4.

Both models also provide an understanding of organ-relative appearance that is consistent with anatomy. In Chapter 3, the profile types correspond to the anatomical context of the kidney, which is surrounded mostly by darker fat but with brighter objects such as spleen, liver, and bones within the boundary region. The choice of profile type per surface point is also consistent with this context. In Section 4.6, the local RIQF cluster populations for the bladder exterior correspond to its anatomical neighbors, the prostate, pubic bone, and bowel, with an additional population representing the mixture of these tissues. These examples show that clustering on local image descriptors, while providing the basis for effective image matches, can also lead to an intuitive understanding of object-relative appearance.

2. *I present a novel appearance model and image match using clustering on intensity profiles that reflects the various edge characteristics in the image boundary region. This model leads to improved kidney segmentation in CT versus a previous image match.*

   The clustered profile model was presented in Chapter 3. In this scheme, voxel-scale intensity profiles observed at fixed object-relative coordinates were clustered via a correlation metric resulting in profile types that reflect the various intensity edges found around the object (Section 3.3). The image match using these profile types was tested in the context of m-rep based segmentation of the kidney in abdominal CT images and compared against a common image match that uses a constant, high contrast-finding profile type. The intensity profile clustering image match led to an improvement in segmentation accuracy in a majority of test images with significantly lower maximum (Hausdorff) error. Furthermore, the clustered profile match proved more robust, as the results obtained required much less parameter tuning compared with the previous match (Section 3.4.1).

   This image match was conceived as an improvement over another profile-based match that uses a constant profile type per point. While the match did prove more effective, there are several shortcomings. For example, it depends on a voxel-scale correspondence in determining the profile type at a point. Additionally, the resulting template is static, providing no information about the variability in the image boundary region (see Section 4.2). Thus, while the profile clustering proved more effective than the constant template it was compared against, the experimental results should be taken only as validation of the general approach of clustering rather than as indication that the profile clustering match is ideal.

3. *I demonstrate that regional intensity quantile functions (RIQFs) are a more effective local image descriptor than intensity profiles.*

   The regional intensity quantile function (RIQF) developed by Broadhurst is a representation of the intensity distribution in an object-relative region (Section 2.3.1). The RIQF has several advantages over the voxel-scale intensity profile (Claim 2) as a local image descriptor, including less susceptibility to noise, a 3D regional description of intensity, and a reliance on a larger scale and thus more trustworthy geometric correspondence. Additionally, Broadhurst designed a means of constructing a probability distribution on RIQFs that allows for a probabilistic appearance model.

   In an experiment conducted jointly with Broadhurst (Section 4.3, Broadhurst et al. (2006)), I compared the efficacy in segmentation of the profile-based match of Chapter 3 against a probabilistic match based on RIQFs of the global interior and exterior near boundary regions. This experiment on the left kidney in CT bore out the advantages of the RIQF as a local image descriptor, making it an obvious choice over intensity profiles.

4. *I present a novel appearance model and image match using clustering on RIQFs. I compare RIQF-based appearance models at several regional scales on bladder and prostate segmentation in CT.*

   This combined-clustered model of Section 4.6 first determines local region types through Gath-Geva clustering on the local RIQFs observed in training. The region types correspond to the neighboring organs and volumes on the exterior of the object of interest. Each local region is aligned to one of the region types. The combined-clustered model then captures the variability in the appearance of these large-scale regions by combining the associated local RIQFs per training case and performing PCA on the resulting sets of corresponding large-scale RIQFs.

The combined-clustered model considers appearance at an intermediate regional scale, smaller than the global scale (Section 4.4.1) but larger than the the scale of the local-geometric model (Section 4.4.3). In Chapter 4 I compared the segmentation accuracy of all three appearance models on bladder and prostate in CT in the context of day-to-day adaptive radiotherapy for the treatment of prostate cancer (Section 4.6.4). I found that the local-geometric model provided the best overall results and that a method to account for the changing object-relative image context of an organ could be beneficial.

5. *I demonstrate that the object-relative positions of neighboring organs and volumes change across images, leading to a false association of image regions due to the geometry-implied correspondence.*

   Figures 5.1 and 5.2 show the changing object-relative position of volumes neighboring the bladder over multiple images of a single patient. The volumes were determined through Gath-Geva clustering of the local RIQFs.

   The figures show that across images there are correspondences based on geometry and correspondences based on image characteristics; these correspondences are often different. The correspondence provided by the m-rep model, which is the geometric correspondence, is determined relative only to the organ of interest and not those neighboring organs and volumes that affect the external appearance. Thus the kind of external variability shown in the figures is not accounted for by appearance models that rely solely on the geometry, with the exception of those relying on more complicated geometries such as diffeomorphic warps or multi-object m-reps (see Section 5.2). Even then, with a geometrically static appearance model (one that relies on a geometrically static correspondence), regions with non-corresponding intensity characteristics may be grouped together, resulting in a false association of image regions and the introduction of artificial variance to the

appearance models.

6. *I present two novel appearance models and associated image matches that account for the variability in the external conformation of neighboring regions through shifting of the image model on the object boundary. The two models differ in the scale of the regions undergoing shifting. These models often lead to improved segmentations over those achieved using their static counterparts. I validate this by experimentally testing the efficacy in segmentation of these appearance models on bladder and prostate in CT.*

   The shifting-combined-clustered and shifting-local-geometric appearance models are presented in Chapter 5. The first extends the combined-clustered model (Section 4.6) by allowing the local regions to change their region-type allegiance in order to fit the current positions of the large-scale regions modeled (Section 5.3). At training time this leads to tighter RIQF models of the external volumes. At target time the shifting is constrained by the configurations observed during training so that the image match returned reflects knowledge of the regional conformations seen in training.

   The shifting-local-geometric model extends this idea to more local object-relative regions (Section 5.4). The shifting at this scale is more heavily constrained due to the noise in sampling very small regions.

   Both models were tested against their static counterparts – the combined-clustered and local-geometric models respectively – in segmenting bladders and prostates in CT in several within-patient leave-one-day-out studies (Section 5.5). Within this context, segmentations using the shifting matches led to very similar and often improved results. The possibility that shifting may help on test sets with more room for improvement is discussed later in this chapter.

Thesis: *The automatic segmentation of medical images via deformable models benefits from the use of object-relative, regional statistical appearance models. Specifically, clustering on either local or regional object-relative image descriptors can provide an effective understanding of the context of the object of interest in the image. Additionally, the use of shifting to account for the variability in the position of object-relative regions allows for relaxation of the geometric correspondence provided by the deformable model and often leads to improved segmentations.*

In this dissertation I have constructed novel appearance models that use clustering and/or shifting on object-relative local and regional image descriptors for use in deformable model segmentation. Within the text these models have been developed and compared against existing methods and each other. Claims 1, 2 and 4 relate to clustering while Claims 5 and 6 relate to shifting. Claim 1 is that clustering-based appearance can improve segmentation. This is proved constructively by Claims 2 and 4, which detail such appearance models and the experiments performed that show their efficacy. Claim 5 is that exterior regions shift relative to the object of interest across images, which is shown to be true in the motivation of Chapter 5. This leads naturally to idea that a shifting appearance model could provide benefit. Claim 6 details two novel shifting appearance models and their efficacy in segmentation of bladders and prostates. Thus overall, appearance models that use clustering or shifting or both on object-relative local and regional image descriptors can improve deformable model segmentation.

6.2 Future Work

This dissertation is the result of my having followed a number of research directions that became apparent during the course of the research. As with all worthy topics of study there are yet still questions to be answered in determining the value to certain applications and the potentials of object-relative appearance models that use clustering

and/or shifting and consider appearance at regional scale. Here, I discuss a collection of research directions in this vein. Many of the proposed research directions are mentioned in the conclusions of Chapters 3, 4, and 5.

6.2.1 Applications

This dissertation is generally concerned with modeling object context using clustering and shifting. I have directly addressed only the segmentation of kidney, bladder and prostate in CT. There are many other organs whose segmentations suffer from the kind of variability in context that my models account for. For example, segmentations of the hippocampus, caudate and other subcortical structures in MRI suffer from the presence of the bordering ventricle. The strong contrast between the ventricle intensity and the tissues would seem to be an advantage in segmentation. However, the ventricle's variable extent along the boundary of an organ confuses other appearance models that use more than simple edge strength – and edge strength is a poor model in other areas of the organs. I expect shifting of the regional appearance to improve such situations, where a clearly delineable region has variable object-relative extent.

Additionally, in segmenting bladder and prostate, the experiments have all been per-patient leave-one-day-out with testing on the left-out day. But the variability in the neighborhood relationships is more present across patients than within a patient (see Section 5.2). Thus I also expect that the benefit provided by shifting will be more clear in across-patient studies. Furthermore, many segmentation tasks outside of adaptive radiotherapy involve training over many patients, in that there may be only one or two images per patient.

6.2.2 Potential improvements to the shifting models

In Chapter 5 I presented appearance models that allow shifting of the trained object-relative appearance in order to account for the variable image conformations seen in

training – which variability is misrepresented or ignored by previous models. I chose to model a shifting region as simply the union of its constituent parts (more local regions) and neighbor relations between those parts. Shifting was a consequence of the changing allegiance of the more local regions. In this subsection I propose several potential improvements to the shifting already described and also a number of ways to reformulate the shifting itself.

Improvements relying on the current, allegiance approach to shifting

The shifting approach described in Chapter 5 iterates over the local regions or subregions on the boundary of the regional delineation and changes their allegiance if it results in a net improvement in the RIQF fits.

One potential improvement would be a greedy approach to picking the local region or subregion to be considered. In the previous shifting, the magnitude of the improvement was ignored beyond the case of choosing between more than two allegiances for a particular local region. A problem with this approach is that it may happen that a subregion is added to a region because it helps somewhat and that new combined region subsequently rejects another subregion that would have helped even more. In the new approach, the potential improvements over all boundary regions would be computed and the change in allegiance leading to the most improvement then chosen.

Another potential improvement, this one at target time, involves a different initial regional delineation for the shifting-combined-clustered scheme. In the presented approach, the shifting starts at the expected regional delineation. In the new approach, the initial delineation would be the optimal allegiance for each local region based on the Gath-Geva clustering. The shifting would then use the prior to retract from this perhaps unlikely conformation in the vein of Iterative Conditional Modes, where the initial conformation is thought of as a noisy labeling of the subregions (Besag, 1986).

Finally, the issue of shifting can be seen as a graph cut problem. Graph cut algo-

rithms are commonly used for the automatic segmentation of color images, where some feature vector is computed for each pixel, the feature vectors are clustered, and the pixels are labeled according their cluster membership. Graph cuts have a speed advantage over iterative schemes in that many labels can change at once. Recent work has also added spatial coherency to the optimization (Zabih and Kolmogorov, 2004), balancing what might be a weak clustering in the feature space against the spatial coherence of the resulting labeling in the image space. The shifting scheme can be fit to this kind of approach, where the feature vectors are the RIQFs at the local regions and the goal is labeling the local regions, analogous to the pixels. At the time of shifting, this could result in a fast, spatially coherent regional delineation. It will be necessary to integrate the larger-scale RIQF fits with the graph cut, which currently only considers the local feature vectors (Boykov et al., 2001).

Improvements via the reformulation of shifting

The potential improvements listed above all rely on the allegiance approach to shifting presented in Chapter 5. Several elements of the problem lend themselves to this approach: there are local regions and feature vectors (RIQFs) associated with them, and the goal is to label each local region. One problem with the approaches above is their discrete nature, where a number of points at fixed positions can change allegiance to affect the overall conformation. In that the neighboring volumes being modeled may not be perfectly aligned with the discrete shifting I have devised, there may be an advantage to a more continuous deformation of the regions. I will describe two possible reformulations of shifting that would allow a smoother deformation.

One possibility is the use of a continuous warp of the sampling regions on the surface of the object. Initially the local regions or subregions are centered at and anchored to points on the surface which are at fixed geometric coordinates (Section 4.4.2). There are many potential warping techniques which can move these regions relative to the

initial local regions by moving the centers of the regions. These techniques have often been designed to improve the *geometric correspondence* between training cases, variously using thin-plate spines (Zou et al., 2007), spherical wavelets (Nain et al., 2005), particle methods (Cates et al., 2007), and functions on the sphere (Davies et al., 2003; Heimann et al., 2005) in conjunction with an objective function based on entropy, covariance, or other measures. Using the warp to place the local and combined regions into better *image correspondence* will lead to tighter RIQF distributions than the discrete, allegiance version. Additionally, the parameters of the transformations across training cases can be stored to constrain the shifting at target time.

Another possibility for smooth shifting is to use contours on the surface that define the sampling domains of the regions. In this approach, a region boundary is specified by the list of profiles on the edge of the sampling domain for that region. During shifting, an individual profile may change allegiance to expand or shrink the region. This is still technically a discrete approach, but at the level of the voxels being added or subtracted. This approach would require research into shape models of contours on the surface.

These approaches to shifting could more accurately reflect the smoothly changing conformation of regions relative to the object. Both schemes involve additional complexity that must be balanced against the possible improvements.

Given a smooth shifting via one of the above approaches, one might consider a probability distribution on such shifts as a means of constraining the changes in conformation. Such a distribution might be on the parameters of the continuous warp or on the shape model of the contour. This is a continuous analog to the scheme presented in Chapter 5, which used a discrete prior on the allegiance of each local region. The prior in that scheme was set to be uniform in training, reflecting a stance that there is no knowledge of the expected regional conformations.

6.2.3 Finite mixture tissue modeling

Some of the appearance models presented in this dissertation have been motivated by considering the image around the object to be made up of homogeneous volumes (Sections 4.4.2, 4.6). In this subsection I propose a finite mixture-based image match that has a similar goal but considers regional appearance from a different angle. Consider that the collection of intensities in a medical image is the result of the mixture of tissues in that image, where each tissue images to some range of intensities. In this context, objects image as regions of intensity or mixtures of intensity with high spatial correlation in the mixture. The image in the vicinity of an object depends on the object and its possible regional subdivision and also the object's exterior divided into its own regions. Many boundaries of concern in medical images can be modeled as changes in the local tissue mixture. In Section 4.4.2, I used clustering on local RIQFs to find these regions of homogeneous mixture. In this section, I propose an image match based on considering the relative amounts of tissues in these regions using finite mixture modeling.

Background

A finite mixture distribution is a weighted average of a number of component distributions. While the component distributions can be any density functions, the theory of finite mixtures is well established in the case when the components are all elements of the same parametric family, such as the Gaussian (Titterington et al., 1985; McLachlan and Peel, 2000). In the Gaussian situation, the finite mixture density is given by

$$p(i|\Psi) = \sum_{k=1}^{K} \pi_k f_k(i) \tag{6.1}$$

$$f_k(i) = p(i|\Theta_k) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-(i-\mu_k)^2/2\sigma_k^2}. \tag{6.2}$$

In the above, $\Theta_k = [\mu_k, \sigma_k^2]$, which are the parameters completely describing the Gaussian intensity distribution of tissue type k. The mixture is parameterized by $\Psi =$

$[\Theta_1, \cdots, \Theta_K, \Pi]$, with $\Pi = [\pi_1, \cdots, \pi_K]$ being the the tuple of additional weighting parameters and $\sum_{k=1}^{K} \pi_k = 1$; that is, Π is a point in the K-simplex, and due to the constraint it lies in a $(K-1)$-dimensional subspace. Given a histogram of the intensities within a region, the parameters Ψ can be computed via an expectation-maximization (EM) scheme (Hasselblad, 1966, 1969). It is a simpler issue to compute the mixture Π given the component parameters Θ_K.

Many previous image match methods use mixture distributions to model the intensities of tissues of interest (Gerig et al., 1992; Priebe et al., 2006). These generally focus on classifying each voxel in an image. A more region-scale approach is taken in (McLachlan et al., 1996), where given the Gaussian parameters for each tissue, the authors determine the mixture of tissues within larger-than-voxel-scale overlapping regions that cover the image; at each voxel, there is an understanding of the mixture of tissues locally. All of these methods have been applied to the voxel classification of MRI images of the brain into the three brain tissue types (grey matter, white matter, cerebrospinal fluid), as opposed to deformable model segmentation in CT.

The tissue mixture image match

The image match I propose considers object-relative regional probabilities on the mixture tuple Π. In this scheme the parameters for each tissue in the training images are determined in preprocessing using model selection (selecting the number of tissues) and finite mixture estimation (Ray et al., 2008; Figueiredo and Jain, 2002) on the set of training images, yielding the set of tissue distributions $\{f_k, \forall k\}$. Given the fixed state of the tissue components, the issue in the remaining training and at target time is only the local estimation of the mixture of these various tissues.

In training, the intensities within a number of object-relative regions indexed by r in image i are sampled as per the local-geometric models presented in this dissertation (Section 4.4.3), yielding the intensity distribution for each region r, q_{ir}. The regional

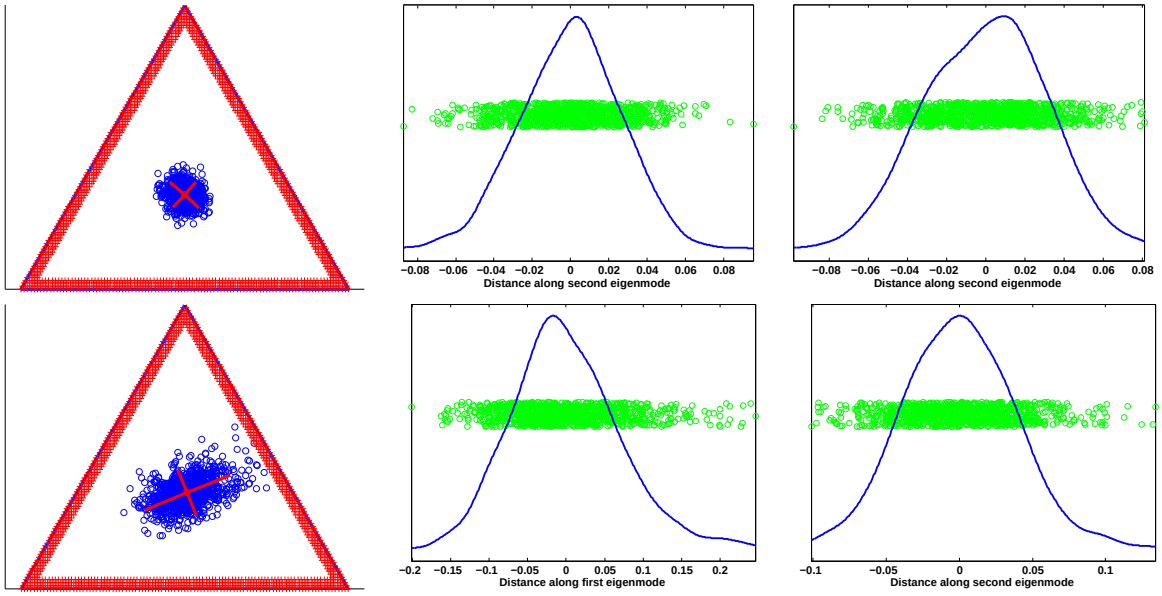

Figure 6.1: Top: samples of the observed mixture tuple Π for some region of three tissue components if the underlying (ideal) mixture is constant. Bottom: the observed mixtures given some variation in the underlying mixture, as might be seen across cases. Across the rows are the tuples plotted in the simplex space and the projection onto the two principal modes. The thick border in the simplex plots show where the multinomial is not well approximated by a Gaussian.

mixture of tissues Π_{ir} is determined using the EM scheme of Hasselblad (1966, 1969) with the simplification that the tissue parameters Θ_K are known. The goal in this match is to model the variability in the set of observed mixtures for the region, $\{\Pi_{ir}, \forall i\}$.

The mixture tuple Π can often be modeled with a Gaussian. If the underlying mixture of tissue within a region is fixed, the variability in the *observed* mixture is due solely to the to the sample size and the relative overlap of the tissue components. The resulting Π observed for a random sample is distributed according to a multinomial, which is approximated by a multivariate Gaussian given the condition that π_k times the sample size is above a heuristic threshold for all k, usually 5 or 10. The covariance structure of the multinomial is given by the inverse of the information matrix. If there are changes in the underlying mixture of the region across training cases, this variability is modeled by a Dirichlet distribution, which is also well approximated by a Gaussian

and thus will appear as additional Gaussian variance. In training then, the mixture model for the region r is given by PCA on the set $\{\Pi_{ir}, \forall i\}$. See Figure 6.1 for an example of PCA on a mixture of three tissue components.

At target time, the image match at a region is given by squared Mahalanobis distance of the observed regional mixture in the PCA space computed for that region. The overall image match sums this match over all regions.

Potential issues

There could be some issues with the tissue mixture match just described. Though the CT intensity of a particular tissue should be standard, there can be differences between images of the same patient across machines, due to calibration error or postprocessing, or between patients in the same machine due to actual differences in the density of particular tissues between patients. This may be addressed through allowing some small change in the underlying tissue components.

An additional concern is partial voluming. The effect of partial voluming is generally a heavy tail in the observed distribution of a particular tissue, with the size of the tail dependent on prevalence of partial voluming effects and the intensity of neighboring tissues. Many have chosen to heuristically model this effect with additional mixture components (Gerig et al., 1992; Priebe et al., 2006), which while convenient can confuse the determination of the underlying intensity distributions on tissue. In contrast, Broadhurst has suggested a principled approach explicitly using the parameters of a Beta distribution within the larger mixture estimation to quantify the amount of partial voluming (Broadhurst, 2008). This extended mixture estimation process could be used within the step of finding Π_{ir} to address the partial voluming problem.

During my research on object-relative appearance models I worked extensively on modeling the intensity distributions within regions using finite mixtures of Gaussians or Gammas. In addition to the issues just described, there are two particular difficulties

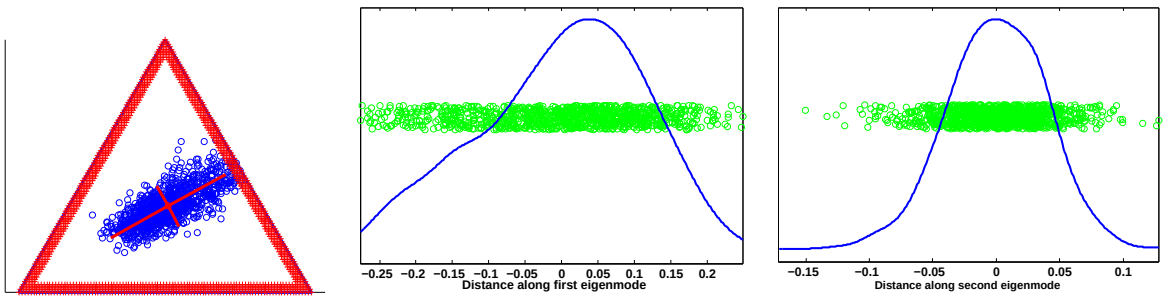

Figure 6.2: Showing PCA on the mixtures when the underlying mixture varies substantially, as is common. Note the heavy tail in the first principal component.

that I found. First, the heuristic conditions on sampling and tissue proportion were often not met for the Gaussianity of the multinomial on Π. A particular tissue component in corresponding regions may be present in large quantity in some cases and virtually absent in others. As the mixture approaches the border region of the simplex, the projections onto the principal directions become decidedly non-Gaussian (see Figure 6.2). Second, the variability in the intensity distribution observed for a given region often involved more than a change in the mixture. The resulting mean and variance changes in the underlying tissue components lead to additional non-Gaussianity in the mixture space that assumes constant tissue component distributions.

However, these difficulties do not indicate that object-relative tissue mixture modeling is hopeless. The difficulties only indicate that a *Gaussian* poorly characterizes the variability in the observed mixture. In future research in this direction, the initial step is to model $p(\Pi)$ as a Dirichlet distribution composed with a multinomial rather than as a Gaussian. There may also be the need to include parameters for changes in the underlying tissue components, resulting in a model for $p(\Psi)$.

6.3 Conclusion

In this dissertation I have proposed novel image appearance models that use clustering and/or shifting in automatic medical image segmentation via deformable models. These

models have been evaluated by considering their efficacy in segmentation of the kidney, bladder, and prostate in abdominal and male pelvis CT. Results show that clustering and shifting on regional image descriptors are promising concepts whose use leads to improved segmentation quality.

Bibliography

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B*, 48(3):259–302. 128

Bezdec, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York. 67, 76

Boykov, Y., Veksler, O., and Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239. 129

Broadhurst, R. E. (2008). *Compact Appearance in Object Populations Using Quantile Function Based Distribution Families*. PhD thesis, the University of North Carolina, Chapel Hill, NC. 24, 28, 55, 57, 81, 134

Broadhurst, R. E., Stough, J., Pizer, S. M., and Chaney, E. L. (2005). Histogram statistics of local model-relative image regions. In *DSSCV Fogh Olsen et al. (2005)*. 28, 60

Broadhurst, R. E., Stough, J., Pizer, S. M., and Chaney, E. L. (2006). A statistical appearance model based on intensity quantile histograms. In *IEEE International Symposium on Biomedical Imaging (ISBI)*. 7, 28, 30, 33, 58, 59, 61, 123

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 8(6):679–697. 14, 41

Caselles, V., Kimmel, R., and Sapiro, G. (1997). Geodesic active contours. *International Journal of Computer Vision*, 22:61–79. 14

Cates, J. E., Fletcher, P. T., Styner, M. A., Shenton, M. E., and Whitaker, R. T. (2007). Shape modeling and analysis with entropy-based particle systems. In *Proc. Int. Conf. on Information Processing in Medical Imaging (IPMI)*, volume 4584 of *Lecture Notes in Computer Science*, pages 333–345. Springer. 9, 130

Chan, T. F. and Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277. 23, 26, 57

Chaney, E. L. (2004). Planunc project website: http://planunc.radonc.unc.edu/. plunc. 49, 61

Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 17(8):790–799. 76

Cootes, T., Beeston, C., Edwards, G., and Taylor, C. (1999). A unified framework for atlas matching using active appearance models. In *Proc. Int. Conf. on Information Processing in Medical Imaging (IPMI)*, volume 1613 of *Lecture Notes in Computer Science*. Springer. 41

Cootes, T., Cooper, D., Taylor, C., and Graham, J. (1995). Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59. 14, 41, 56

Cootes, T., Edwards, G., and Taylor, C. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685. 19

Cootes, T. F., Hill, A., Taylor, C. J., and Haslam, J. (1994). Use of active shape models for locating structures in medical images. *Image and Vision Computing*, 12(6):355–365. 17

Costa, M. J., Delingette, H., and Ayache, N. (2007). Automatic segmentation of the bladder using deformable models. In *IEEE International Symposium on Biomedical Imaging (ISBI)*. 26, 27, 37

Davies, R. H., Twining, C. J., Allen, P. D., Cootes, T. F., and Taylor, C. J. (2003). Shape discrimination in the hippocampus using an mdl model. In *Proc. Int. Conf. on Information Processing in Medical Imaging (IPMI)*, volume 2732 of *Lecture Notes in Computer Science*, pages 38–50. Springer. 9, 130

Davis, B. (2008). *Medical Image Analysis via Frechet Means of Diffeomorphisms [and Manifold Kernel Regression]*. PhD thesis, the University of North Carolina, Chapel Hill, NC. 16, 95

Davis, B., Fletcher, P. T., Bullitt, E., and Joshi, S. (2007). Population shape regression from random design data. In *International Conference on Computer Vision*. 16

Dice, L. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26:297–302. 38, 73

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern Classification*. Wiley Interscience. 20, 24

Duta, N., Jain, A. K., and Dubuisson-Jolly, M.-P. (1999). Learning-based object detection in cardiac mr images. In *ICCV*, pages 1210–1216. 18

Eberly, D. (1999). Distance between point and triangle in 3d. found at http://www.geometrictools.com/Documentation/DistancePoint3Triangle3.pdf. 37

Fenster, S. D. and Kender, J. R. (2001). Sectored snakes: Evaluating learned-energy segmentations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9):1028–1034. 18, 41

Figueiredo, M. A. T. and Jain, A. K. (2002). Unsupervised learning of finite mixture models. *IEEE Transactions Pattern Analysis Machine Intelligence*, 24(3):381–396. 132

Fletcher, P., Lu, C., Pizer, S., and Joshi, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005. 35, 71

Fogh Olsen, O., Florack, L., and Kuijper, A., editors (2005). *Deep Structure, Singularities and Computer Vision*, volume 3753 of *Lecture Notes in Computer Science*. Springer–Verlag. 137, 141

Freedman, D., Radke, R., Zhang, T., Jeong, Y., Lovelock, D., and Chen, G. (2005). Model-based segmentation of medical imagery by matching distributions. *IEEE Transactions on Medical Imaging*, 24(3):281–292. 26, 27, 28, 37, 57

Gan, R. and Chung, A. C. S. (2005). Multi-dimensional mutual information based robust image registration using maximum distance-gradient-magnitude. In *IPMI*, pages 210–221. 17

Gath, I. and Geva, A. (1989). Unsupervised optimal fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):773–780. 76

Gerig, G., Jomier, M., and Chakos, M. (2001). Valmet: A new validation tool for assessing and improving 3d object segmentation. In Niessen, W. and Viergever, M., editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 2208, pages 516–523, New York. Springer. 36, 38

Gerig, G., Martin, J., Kikinis, R., Kübler, O., Shenton, M. E., and Jolesz, F. A. (1992). Unsupervised tissue type segmentation of 3d dual-echo mr head data. *Image and Vision Computing*, 10(6):349–360. 20, 132, 134

Han, Q., Merck, D., Levy, J., Villarruel, C., Damon, J. N., Chaney, E. L., and Pizer, S. M. (2007). Geometrically proper models in statistical training. In *Proceedings, Information Processing in Medical Imaging (IPMI)*, volume 4584 of *Lecture Notes in Computer Science*, pages 751–762. 34, 115

Han, Q., Pizer, S. M., Merck, D., Joshi, S. C., and Jeong, J.-Y. (2005). Multi-figure anatomical objects for shape statistics. In *Proceedings, Information Processing in Medical Imaging (IPMI)*, volume 3565 of *Lecture Notes in Computer Science*, pages 701–712. 33

Hasselblad, V. (1966). Estimation of parameters for a mixture of normal distributions. *Technometrics*, 8(3):431–444. EM for grouped (histogram) data. 132, 133

Hasselblad, V. (1969). Estimation of finite mixtures of distributions from the exponential family. *Journal of the American Statistical Association*, 64:1459–1471. EM for grouped (histogram) data, not just Gaussians. 132, 133

Heimann, T., Wolf, I., Williams, T. G., and Meinzer, H.-P. (2005). 3d active shape models using gradient descent optimization of description length. In *Proc. Int. Conf. on Information Processing in Medical Imaging*, volume 3565 of *Lecture Notes in Computer Science*, pages 566–577. Springer. 130

Ho, S. (2004). *Profile Scale Spaces for Statistical Image Match in Bayesian Segmentation*. PhD thesis, the University of North Carolina, Chapel Hill, NC. 25, 54

Jain, A. K., Duin, R. P., and Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37. 20

Joshi, S., Pizer, S., Fletcher, T., Thall, A., and Tracton, G. (2001). Multi-scale 3-d deformable model segmentation based on medial description. In *Proc. Int. Conf. on Information Processing in Medical Imaging (IPMI)*, volume 2082 of *Lecture Notes in Computer Science*, pages 64–77. Springer. 15, 33

Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: active contour models. *International Journal of Computer Vision*, 1(4):321–331. 14

Kelemen, A., Székely, G., and Gerig, G. (1999). Elastic model-based segmentation of 3-d neuroradiological data sets. *IEEE Transactions on Medical Imaging*, 18(10):828–839. 18, 25

Leventon, M., Faugeras, O., Grimson, W., and Wells, W.M., I. (2000). Level set based segmentation with intensity and curvature priors. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA)*, pages 4–11. 24

Levina, E. (2002). *Statistical Issues in Texture Analysis*. PhD thesis, University of California at Berkeley. 28

Lorensen, W. E. and Cline, H. E. (1987). Marching cubes: A high resolution 3d surface construction algorithm. In *SIGGRAPH 87*, volume 21, pages 163–169. 36

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997). Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198. 16

Mahalanobis, P. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science of India*, 12:49–55. 5

Maintz, J. and Viergever, M. (1998). A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36. 22

McLachlan, G., Ng, S., Galloway, G., and Wang, D. (1996). Clustering of magnetic resonance images. *Proc. ASA (Statistical Computing Section)*, pages 12–17. 20, 132

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Wiley Interscience. 131

Merck, D., Tracton, G., Saboo, R., Levy, J., Chaney, E., Pizer, S., and Joshi, S. (2008). Training models of anatomic shape variability (to appear). *Medical Physics*. 34, 95, 115, 116

Miller, M., Joshi, S., and Christensen, G. (1999). *Brain Warping, Ed: Arthur Toga*, chapter Large Deformation Fluid Diffeomorphisms for Landmark and Image Matching, pages 115–132. Elsevier Science & Technology Books. Academic Press. 15

Montagnat, J. and Delingette, H. (1998). Globally constrained deformable models for 3d object reconstruction. *Signal Processing*, 71:173–186. 41

Nain, D., Haker, S., Bobick, A., and Tannenbaum, A. (2005). Multiscale 3d shape analysis using spherical wavelets. In *Proceedings, Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 3750 of *Lecture Notes in Computer Science*, pages 459–467. 130

Niessen, W., Bouma, C., Vincken, K., and Viergever, M. (2000). *Performance Characterization in Computer Vision*, chapter Error Metrics for Quantitative Evaluation of Medical Image Segmentation, pages 275–284. Kluwer Academic Publishers (Springer). 36

Pekar, V., McNutt, T. R., and Kaus, M. R. (2004). Automated model-based organ delineation for radiotherapy planning in prostatic region. *International Journal of Radiation Oncology, Biology, Physics*, 60(3):973–980. 37

Pizer, S., Fletcher, T., Fridman, Y., Fritsch, D., Gash, A., Glotzer, J., Joshi, S., Thall, A., Tracton, G., Yushkevich, P., and Chaney, E. (2003). Deformable m-reps for 3d medical image segmentation. *International Journal of Computer Vision - Special UNC-MIDAG issue*, 55(2):85–106. 15, 17, 33, 71

Pizer, S. M., Fletcher, P. T., Joshi, S., Gash, A. G., Stough, J., Thall, A., Tracton, G., and Chaney, E. L. (2005a). A method & software for segmentation of anatomic object ensembles by deformable m-reps. *Medical Physics Journal*, 32(5):1335–1345. 33, 44

Pizer, S. M., Jeong, J., Lu, C., Muller, K., and Joshi, S. (2005b). Estimating the statistics of multi-object anatomic geometry using inter-object relationships. In *DSSCVFogh Olsen et al. (2005)*. 95

Pratt, W. K. (1991). *Digital Image Processing*. Wiley, New York. 22

Priebe, C., Miller, M., and Ratnanather, J. (2006). Segmenting magnetic resonance images via hierarchical mixture modelling. *Computational Statistics and Data Analysis*, 50:551–567. 20, 132, 134

Rao, M., Stough, J., Chi, Y.-Y., Muller, K., Tracton, G., Pizer, S. M., and Chaney, E. L. (2005). Comparison of human and automatic segmentations of kidneys from ct images. *International Journal of Radiation Oncology, Biology, Physics*, 61(3):954–960. 41, 54

Ray, S., Lindsay, B. G., and Li, J. (2008). Modal em for mixtures and its application in clustering. 132

Rousson, M., Khamene, A., Diallo, M., Celi, J., and Sauer, F. (2005). Constrained surface evolutions for prostate and bladder segmentation in ct images. In *CVBIA*, volume 3765 of *Lecture Notes in Computer Science*, pages 251–260. Springer. 95

Rubner, Y., Puzicha, J., Tomasi, C., and Buhmann, J. M. (2001). Empirical evaluation of dissimilarity measures for color and texture. *Computer Vision and Image Understanding*, 84:25–43. 24, 28, 57

Scott, I., Cootes, T., and Taylor, C. J. (2003). Improving appearance model matching using local image structure. In *Proc. Int. Conf. on Information Processing in Medical Imaging (IPMI)*, volume 2732 of *Lecture Notes in Computer Science*, pages 258–269. Springer. 20

Shewchuk, J. R. (1994). An introduction to the conjugate gradient method without the agonizing pain. 5

Staib, L. H. and Duncan, J. S. (1992). Boundary finding with parametrically deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(11):1061–1075. 14

Stough, J. V., Broadhurst, R. E., Pizer, S. M., and Chaney, E. L. (2007). Clustering on local appearance for deformable model segmentation. In *IEEE International Symposium on Biomedical Imaging (ISBI)*. 66

Studholme, C., Hill, D. L. G., and Hawkes, D. J. (1999). An overlap invariant entropy measure of 3d medical image alignment. *Pattern Recognition*, 32(1):71–86. 17

Su, T. and Dy, J. (2004). A deterministic method for initializing k-means clustering. In *IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*. 77

Tewell, M. A. and Adams, R. (2004). The plunc 3d treatment planning system: a dynamic alternative to commercially available systems. *Medical Dosimetry*, 29(2):134–138. 49

Titterington, D., Smith, A., and Makov, U. (1985). *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons. Sec. 4.3.2 for EM. 131

Tsai, A., Anthony Yezzi, J., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W. E., and Willsky, A. (2003). A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Transactions on Medical Imaging*, 22(2):137–154. 23, 24, 26, 27, 57

Tsai, A., Wells, W., Tempany, C., Grimson, W. E., and Willsky, A. (2004). Mutual information in coupled multi-shape model for medical image segmentation. *Medical Image Analysis (MedIA)*, 8:429–445. 17, 95

Turk, G. and O'Brien, J. F. (1999). Shape transformation using variational implicit functions. In *SIGGRAPH*, pages 335–342. 27

Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., and Smith, S. M. (2005). Mixture models with adaptive spatial regularization for segmentation with an application to fmri data. *IEEE Transactions on Medical Imaging*, 24(1):1–11. 20

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., and Gerig, G. (2006). User-guided 3d active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage, vol. 31*, 31(3):1116–1128. 14, 23

Zabih, R. and Kolmogorov, V. (2004). Spatially coherent clustering using graph cuts. In *Computer Vision and Pattern Recognition (CVPR)*. 129

Zhu, S. C. and Yuille, A. (1996). Region competition: Unifying snakes, region growing, and bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900. 22, 57

Zitová, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000. 22

Zou, G., Hua, J., and Muzik, O. (2007). Non-rigid surface registration using spherical thin-plate splines. In *Proceedings, Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 4791 of *Lecture Notes in Computer Science*, pages 367–374. 119, 130